

Affective L2 learning experiences
and ideal L2 selves
in spoken CALL practice

This work is part of the research program Feedback and the Acquisition of Syntax in Oral Proficiency (FASOP) NWO-Dossier Number 360-75-010, which was funded by the Netherlands Organisation for Scientific Research (NWO).

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111
e-mail: lot@uu.nl
<http://www.lotschool.nl>

Cover illustration: Ode to Melvin, by Sophie Ripp and Stephen Bodnar

ISBN: 978-94-6093-210-6
NUR: 616

Copyright © 2016 Stephen Bodnar. All rights reserved.

Affective L2 learning experiences
and ideal L2 selves
in spoken CALL practice

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 30 juni 2016
om 12.30 uur precies

door

Stephen Bodnar

geboren op 24 maart 1978
te Richmond Hill (Canada)

Promotor: Prof. dr. R.W.N.M. van Hout

Copromotoren: Dr. C. Cucchiarini
Dr. W.A.J. Strik

Manuscriptcommissie:
Prof. dr. J.J.M. Schoonen
Dr. H. Strating
Prof. dr. P. Desmet (KU Leuven, België)

Contents

Acknowledgements	vii
1 Introduction	1
1.1 Affective dimensions in language learning	2
1.2 Computer-assisted language learning	8
1.3 GREET: A platform for researching affective dimensions in spoken CALL practice	14
1.4 Goals and outline of this thesis	17
2 Evaluating the motivational impact of CALL systems: current practices and future directions	19
2.1 Introduction	21
2.2 Perspectives on L2 motivation	22
2.3 From theory to CALL practice: key ideas for evaluating motiva- tional impact	26
2.4 Evaluating the motivational impact of CALL systems	28
2.5 Future Directions	42
2.6 Conclusion	49
3 Learner affect in computerised L2 oral grammar practice with corrective feedback	51
3.1 Introduction	53
3.2 Background	54
3.3 The present study	56
3.4 Methods	57
3.5 Results	62
3.6 Discussion	66
3.7 Pedagogical implications	68

4 Computerised oral L2 grammar self-study: affective experience, corrective feedback, initial proficiency and learning gain	71
4.1 Introduction	73
4.2 Research background	74
4.3 The present study	76
4.4 Materials and methods	77
4.5 Results Part one: Affective experiences in CALL-ASR grammar practice and learning gains	83
4.6 Results Part two: Relating in-practice affective experiences to initial proficiency and learning gains	87
4.7 Discussion and conclusions	93
5 Investigating motivational diversity through a possible selves approach: a computer application for adult L2 learners of Dutch	97
5.1 Introduction	99
5.2 Background	101
5.3 A computer activity to elicit information concerning L2-related self-guides	108
5.4 Using the activity to investigate the diversity of Dutch learners' ideal L2 selves	117
5.5 Results	123
5.6 General Discussion	131
5.7 Conclusions and future directions	138
5.8 Appendix	139
6 Discussion and conclusions	145
6.1 Significance of the experimental platform developed	147
6.2 What factors influence in-practice affect in ASR-CALL self-study?	149
6.3 The contribution of behaviour logs to understanding affective states	155
6.4 Profiles and prevalence of Dutch learners' ideal L2 selves	157
6.5 General conclusions	159
6.6 Limitations	160
6.7 Future directions	161
6.8 Final words	164
References	165
Contributions	185
Samenvatting in het Nederlands	189
Curriculum Vitae	197

Acknowledgements

The work I completed during my doctorate studies, culminating in this thesis, would not have been possible without my supervisors. Roeland, Catia and Helmer, there are many reasons to thank you. What comes to mind first is the freedom you gave me to explore areas of research that I found interesting but were not part of the original goals for the project. I'm also immensely grateful for your constant availability. Whether on weekdays, evenings, or weekends, you always responded quickly. Each of you contributed something different to the group. Roeland, your guidance during data analysis was of great help. The pressure you applied and your involvement at the end of the project was important in keeping me motivated. Helmer, with the structure you brought to our work, you made sure that everything stayed on track, in meetings, writing, and planning (!). I also appreciated your understanding of the software engineering that went into building the system. Catia, your input during our group discussions, feedback during planning, and advice during the writing of these chapters were invaluable. Your editing super-powers pulled me out of a jam more than once!

During my project I worked in close collaboration with another PhD candidate. Bart, not only were you an excellent colleague to have but also a good friend with a great sense of humour. I really appreciated the many coffee breaks and 8.11 brainstorming sessions. It was great having you to discuss on- and off-topic literature, help test the software, provide feedback on papers, share running the experiments, and the list goes on. I hope we have another opportunity to collaborate.

A number of other people also made significant contributions to this thesis. Thanks to Joost van Doremalen for developing the ASR module in GREET, and later setting up the ASR as a web service, as well as for help administrating the server. For help annotating the many utterances we collected and trying out different research ideas with the GREET platform, I would like to thank Susan and the great students I was lucky enough to work with: Polina, Irene, Kimberly, Judith, and Kimber. In addition, I am grateful to the reading committee for evaluating my thesis and providing helpful comments.

The experiments described in this thesis were made possible with help from a number of educational centers teaching Dutch as a second language. I would like to thank José Bakx and the Dutch teachers at Radboud In'to Languages for always being interested in our research and allowing us to visit their classrooms to recruit participants. Special thanks to Petra Roël for reviewing and providing feedback on the material in the spoken grammar practice, and to Vivika Jagtenborg for test-driving and providing feedback on the GREET platform. At the HAN, I would like to thank Manja Mekking for advising us on which students to target and providing an opportunity to gather participants. A special thanks also goes out to Marijke Bakkeren and Marijke Bleeker of the MMS, who generously allowed us to use the center's excellent computer lab facilities for our experiments and were always very friendly and helpful. I also thank Jeanette Verdonk at I & I Training en Opleiding for allowing me to pilot my motivation activity with lower-educated learners. Thanks also to Ineke van de Craats for her help with creating a Dutch version of the motivation activity and extending the grammar practice material, and to Henk who kindly let Bart and I use the SPEX lab on the eighth floor for our experiments.

I would like to thank the colleagues I worked with on the eighth floor of the Erasmus building: Eva, Christina, Mario, Marjoke, Christel, Sara, Michele, Remy, Odette, Inge, Richard, Mira, Marijn, Maarten, Anna, Wessel, Matyas, Dirkje, Nelleke and Hella. You are a great bunch of people and I feel lucky to have been able to share coffee breaks and drinks at the Cultuur Café with you. Special thanks go to Eric, for being one of my paranymphs, a road warrior, and an all-around great guy, and to my former officemates Louis and Vanja (deep bow) and Sun Yang and his cup (vase) of water.

Besides having great colleagues, I also had amazing friends who made Nijmegen a very enjoyable place to live. Else, thank you for a wonderful introduction to the Netherlands. Huge thanks also to Francisco, Jeremy and Kat, Job and Bonnie, Harald, Simeon and Rosa, Sarah T.H., Folkert and Marloes, Esther, Jana, Wencui, Josh, and Bob. Hartstikke bedankt to Pascal and Ellen for translating my summary into Dutch, and for awesome summer evenings in Heumen.

In Montreal, I would like to thank Prof. Susanne Lajoie and Phillipe Latour, as well as everyone at the McGill ATLAS Lab, for their support and encouragement during the final stages of thesis writing.

In my rush to submit my Master's thesis, I forgot to include a page with acknowledgements and that has always bugged me. I would like to take this chance to thank Dr. Robert Dale for supervising my research project, and Dr. Robert Mannell for coordinating the programme.

On a personal note, I thank my mother and Tilman, for their emotional support, and my father, Mike, Bronwyn, Mark and Viena for encouragement during the last few months. Finally, my thanks go to Sophie, who supported me in every way possible while working towards this thesis.

CHAPTER 1

Introduction

Learning a foreign or second language (L2) is a long process that can begin for different reasons. People may find it intrinsically ‘cool’, view it as practical career strategy, or have an admired role model (Thompson & Vásquez, 2015); others may begin as a result of immigration or asylum, where they have less of a real choice. Regardless of how they begin and the paths on the way to learning a language, it is clear that learners will encounter challenges and that, contrary to what is sometimes claimed on the internet, most will need to sustain their effort to learn the language over a long period of time, through periods of development and stagnation, and in a variety of learning situations. Research into individual differences in second language acquisition (SLA) has aimed at an understanding of why some language learners are more succesful than others (Ellis, 1994), with one line of inquiry investigating the role of an individual’s personality, motivation, attitudes, beliefs, values, goals, expectations, perceptions and emotions, together referred to as affective characteristics. Research into these areas over the past two decades has led to a number of exciting theoretical and methodological developments, including but not limited to increased attention to the language learner’s self-concept, how affective characteristics change with time, and the context in which a person learns and with which they interact.

Over the years, technology and computer-based learning environments have become an important context for language learning. The study of language learning in computerised environments is known as Computer-Assisted Language Learning (CALL). CALL is a broad field which can involve a variety of different technologies and settings, including online distance and blended learning, language learning with mobile technology, and self-study

computer applications that use speech or language technology to support natural language interactions. CALL is also a field with a diverse set of stakeholders, including learners, teachers, CALL system developers, researchers and commercial enterprises, each holding a different perspective. As a field, the emphasis in CALL seems to be on practical issues related to language learning, such as pedagogy and design, although there is also a clear recognition of the importance of research (e.g. Levy, Hubbard, Stockwell, & Colpaert, 2015). Research in CALL can address a broad variety of topics, but speaking generally, might be classified into two broad areas: technological innovations, and pedagogical effectiveness and an understanding of the learning process.

The research presented in this thesis concerns the combination of two areas, affective factors influencing language learning and research into pedagogical effectiveness and learning processes, in language learning that takes place in computerised contexts. Specifically, we examine a group of L2 Dutch learners' affective experiences during language practice, together with factors that are thought to influence the experience, in the context of a self-study computer application providing oral L2 practice targeting Dutch syntax. Our approach involves two different perspectives: 1) a situated perspective which focuses on how learners' affective states change with time and the relationship between affect, practice behaviour and learning outcomes, and 2) a global perspective, which investigates the contents and diversity of learners' L2-related hopes for the future by employing an innovative web-based card sorting activity.

The remainder of this chapter is organised as follows. In section 1.1, we provide a brief overview of research into affective factors implicated in the L2 learning process. Section 1.2 introduces research in Computer-assisted Language Learning (CALL), with attention to computer applications that provide spoken language practice. This is then followed in section 1.3 by an overview of an experimental CALL platform we developed which formed the practice context for our research. In section 1.4, we present our research aims and outline the structure of this thesis.

1.1 Affective dimensions in language learning

What is meant by 'affective dimensions' of language learning? In research aimed at understanding how people learn a second language, one common direction has been to search for and understand cognitive factors that might be involved. Research in this direction contributes to our understanding of second language acquisition (SLA) and to what makes for an effective learning environment. At the same time, from an educational point of view, an account of language learning in terms of cognitive processes alone is insufficient for explaining the differences we routinely observe in the classroom, and the extent to which learners vary in paying attention, exerting effort, and enjoying learning. As educators with an interest in seeing L2 learners progress, we seek the ability to positively influence these aspects by designing or manipulating learning

situations. It therefore becomes important to explain why we observe these variations, and for this an understanding of affective dimensions of learning is necessary, i.e. dimensions of learning related to the learner's subjective perceptions, attitudes, motivations and emotions. Affective characteristics such as these are central to the research presented in this thesis. Below, we provide an overview of related research in L2 motivation (interested readers may wish to refer to more comprehensive treatments provided by Dewaele, 2009; Ushioda & Dörnyei, 2012).

1.1.1 Social-psychological perspectives

Research into affective factors influencing language learning began with a quantitative, social-psychological approach to examining how learners' attitudes related to success in acquiring a particular L2. This approach to studying affective dimensions of L2 learning is exemplified by the work of Gardner and associates (e.g. Gardner, 1985b; Gardner & Lambert, 1959), who are recognised for pioneering the study of L2 motivation (Dörnyei & Ushioda, 2011, p. 40).

An important contribution of Gardner's work is the influential Attitude / Motivation Test Battery (AMTB) (Gardner, 1985a) which consists of a number of multi-item scales designed to measure affective constructs related to L2 learning. Examples of the type of L2-related affective characteristics elicited by the questionnaire include learner's attitudes towards social interaction with the community of native speakers, the potential pragmatic benefits of learning an L2 or attitudes and beliefs concerning their learning environment (specifically, their L2 teacher and course). Questionnaire data is processed using factor analysis to indicate which scales can be grouped together. Following this, learners' motivational attributes are obtained by computing a number of composite indices, e.g. learners' *integrativeness*, defined as their desire to learn the L2 for "integrative (or social) reasons, and his/her general interest in other languages" (Gardner, 1985a, p. 4), and learners' attitudes towards their learning situation. The point is that, for research in the socio-psychological tradition, data was obtained from a single administration of a questionnaire which elicited information from a large number of students' on their affective characteristics at a general or global level.

For various reasons, researchers began to voice criticisms regarding the limitations of this approach to studying affective dimensions of language learners. The one most relevant to this thesis can be attributed to an influential review by Crookes and Schmidt (1991) which argued that Gardner's work and other work in the socio-psychological tradition was not particularly applicable to educators. The problem can be traced to the level of detail Gardner chose to work. In choosing relatively stable affective characteristics (perhaps relatively unmalleable), Gardner's work can be seen to take a macro-perspective that, as Ushioda (1994) has stated, can be described as zooming out from the details of the processes and effectively taking a "snapshot" (p. 81) of the affective characteristics and their relation to learning behaviours and L2

achievement (see also Macintyre, 2007). It should be clear that this type of knowledge is not readily applicable to educator concerns, which has much more to do with students' attention, effort and enjoyment in particular language learning environments, in or outside the classroom. Later, we will return to macro-perspective L2 motivation approaches to describe recent developments, but before that we look at another trend, namely situated perspectives that emerged during the 1990s.

1.1.2 Development of situated perspectives

In response to these critiques the field entered the 1990s with a “mood for expansion and integration” (Dörnyei & Ushioda, 2011, p. 47). In Dörnyei's view (2005), the developments that occurred during this period are best captured in two trends: the integration of cognitive motivational concepts and the development of situated perspectives that involve “a more fine-tuned and situated analysis of motivation as it operates in actual learning situations” (Dörnyei, 2005, p. 74). Of these, the second trend, situated perspectives, is most relevant here.

Situated perspectives can be seen as attempts to move the field towards a more educator-validated perspective on affective dimensions on language learning. This involved re-focusing from the macro-perspective taken in socio-psychological approaches to zoom in to relatively situation-specific or situation-connected affective dimensions. For example, Crookes and Schmidt (1991) advocated investigating dimensions such as the interest, satisfaction, relevance and value a learner experienced within a particular learning environment. Affective dimensions conceptually more similar to emotions (e.g. ‘feelings’) like anxiety emerge (Gardner, Day, & MacIntyre, 1992; Gardner & MacIntyre, 1993). Later, the notion of state motivation is proposed, referring to a learner's “motivational condition at a particular point in time” (Tremblay, Goldberg, & Gardner, 1995, p. 358), and therefore applicable to affective experiences in real L2 practice situations such as the classroom or computer lab. A good example of a situated approach is the study by Julkunen (2001) which examined how state motivation was influenced by different tasks and different learning situations. This research drew on Boekaert's model of motivation (1988) which includes motivational traits, motivational state and task appraisal. The study had an experimental design that sampled motivational state using a short questionnaire before and after different vocabulary tasks, in various learning situations (individual learning, group learning and competitive learning).

1.1.3 Increasing attention to context and time

Another important change, which took place towards the end of the decade, was an increase in attention to the temporal dimension of motivation. As Ushioda (1996) noted, Gardner and MacIntyre (1993) recognised that the “old characterization of motivation in terms of integrative vs. instrumental

orientations is too static and restricted” (p. 4) and subsequently presented an adapted socio-education model to reflect the idea that motivation is both an engine that fuels language learning, and an outcome that is shaped by past or present learning experiences. A model that articulated the temporal aspect of motivation was developed by Williams and Burden (1997) who proposed a model of motivation that included an explicit time dimension, differentiating between three phases: 1) the period before the opportunity to engage in an activity is encountered, in which global attitudes and other trait-like dispositions are typically studied; 2) the period in which the learner is deciding whether or not to undertake an activity; 3) the period when the learner, already engaged in the activity, decides whether or not to sustain their activity. Another model proposed during this period by Dörnyei and Ottó (1998) was the process model, which provided a detailed description of motivation related cognitive processes across time (see also Dörnyei, 2000). It hypothesizes three phases of motivated behaviour and specifies relevant motivational influences that “underlie and fuel the behavioural process” (Dörnyei & Ushioda, 2011, p. 65). Very briefly, in the pre-actional phase “initial wishes, hopes and desires are first transformed into goals, then into intentions” (p.65). The actional phase corresponds to when the learner is “actually embarking on the task, ... [and] the individual is committed to action” (p. 65). Finally, the post-actional phase involves “critical retrospection after action has been completed or possibly interrupted for a period” (p. 66).

In an approach distinct from the work described above, studies by Ushioda (1993, 1994); Ushioda (1996) can be said to develop a learner-validated approach to the study of affective dimensions in language learning. Ushioda’s work is interesting because, in contrast to previous theory-based models, it can be seen to have a data-driven approach to researching motivation. Her work is also noteworthy for developing the temporal aspect of motivation and associated affective dimensions of language learning. She acknowledges the value of Crookes and Schmidt’s (1991) suggestion to adopt an educator-validated approach to investigating affective dimensions of learning, but astutely points out that a learner-validated concept is at least equally as valuable (Ushioda, 1993). In a departure from the social-psychological quantitative tradition, Ushioda’s research (1993; 1994; 1996) explores learners’ sources of motivation qualitatively, by eliciting spontaneous explanations of motivation from learners in open and semi-structured interviews over a 16-month period. Analyses from the study revealed new sources of L2 motivation: in addition to being motivated by future goals to live abroad, learners also frequently attributed their motivation to positive feelings of competence and ability, based on past learning experiences, and to their enjoyment of language learning (Ushioda, 1993, 1994). Second, the study revealed a number of qualitatively different patterns of change, such as short-term incentives overriding more typical motivational characteristics, changes in L2-related personal goals, and changes due to non-L2 related factors (Ushioda, 1996). Ushioda (1996) finds that her observations are not accurately described by existing models of L2 motivation

produced in quantitative approaches and interprets this to support her view that a qualitative approach may be more suitable.

In sum, the studies mentioned in the current section show that past or present perceptions (of competence) or affective experiences (enjoyment) are also important influences on motivation and that, over time, there can be various types of qualitatively different motivational changes.

1.1.4 Modern situated approaches

Modern situated approaches to studying motivation and affect continue to highlight the importance of the temporal dimension. This is particularly evident in a theoretical piece by Ushioda (2009b) articulating a Person-in-Context-relational view of motivation. The view emphasises the role of context and has strong similarities with socio-cultural theory. Important in this view is the acknowledgement that motivation in a situated context can change over time and that motivation emerges from the interaction of the learner and their context. In later work, Ushioda (2012) suggests complementing self-report data with empirical observations of learner behaviour and studies of the relationship between motivation and the events that occur in the learning environment. In Chapter 2 we go into more detail on Ushioda's perspective and the impact this has had on our own approach. Next we return to macro approaches to L2 motivation to examine more recent developments.

1.1.5 Modern global approaches

During the last decade the field's approach to the affective dimension of language learning has also seen important changes. One way the field expanded was by using essentially the same socio-psychological approach but employing constructs from other theories related to affect and motivation. For example, some studies explored studying motivation by employing the constructs of intrinsic and extrinsic motivation from Self-Determination Theory (Noels, 2009). Another example is the conceptualisation of the influential willingness-to-communicate model during the period (MacIntyre, Dörnyei, Clément, & Noels, 1998). Studies such as these are examples of ways the field expanded to include affective dimensions previously unapplied to L2 learning.

A second important development was the proposal of a new theoretical framework by Dörnyei (2005) which addresses a second limitation of Gardner's theoretical framework. This had to do with Gardner's conceptualisation of integrativeness and its applicability to the special case of learners of English as a foreign language (Ushioda, 2011). Briefly, this line of criticism holds that English has become a global international language with no obvious target language community, yet modern learners of English still hope to join the global community of English speakers, yet this 'group' cannot readily be defined in terms of culture or traditions. Thus, there was a need to reconceptualise integrativeness to better handle the case of global English.

The theory is called the L2 Motivational Self System (L2MSS) (Dörnyei, 2005, 2009a). In developing the theory, Dörnyei drew on motivational psychology literature theorising about the motivational and emotional effects of people's thoughts concerning the kind of person they could become in the future. In the related theories (Higgins, 1987; Markus & Nurius, 1986), cognitive representations of people's potential future selves are called *possible selves*. These future thoughts are theorised to have the important property that they have the capacity to induce emotional states or motivate behaviour. In the domain of L2 learning, Dörnyei (2009a) hypothesises that learners' motivation and affective state is linked to two important L2-related future selves. The Ideal L2 self represents the type of L2-speaking person the learner hopes to become, such as a fluent L2 speaker or an L2 speaker with many friends. On the other hand, the Ought-to L2 self represents the kind of L2-speaking person the learner feels obligated to avoid becoming, such as the speaker who fumbles their speech during a public talk, or a person who fails an L2 proficiency test. The third component is the L2 learning experience, which contains "executive" motives related to the immediate learning environment and experience (e.g. the impact of the teacher, the curriculum, the peer group, the experience of success" (Dörnyei, 2009a, p. 47). It is related to the "motivation inspired by prior experience interacting with the present learning environment" (MacIntyre, Mackinnon, & Clement, 2009, p. 68).

To date, the L2MSS has received much attention (Boo, Dörnyei, & Ryan, 2015). Concerning empirical investigations, much research has taken a social-psychological approach employing Ideal and Ought-to L2 scales as predictors of criterion measures, such as intended effort and motivated behaviour, frequently alongside more traditional affective dimensions. This has the benefit of establishing links to compare the reinterpretation of the integrative motive from a possible selves perspective. With a number of large-scale studies providing supporting evidence the validity of the L2MSS (e.g. Al-Shehri, 2009; Csizér & Lukács, 2010; Islam, Lamb, & Chambers, 2013; Kim, 2012; Taguchi, Magid, & Papi, 2009), an important area for future research is to examine the basis for possible selves capacity to motivate (Taylor, 2013). Only a few studies have investigated how different attributes of an L2 self (e.g. how desired they are, how frequently they are thought of, how realistic they are) are related to their potential motivating effects (Hessel, 2015; MacIntyre, Mackinnon, & Clement, 2009).

1.1.6 Affective dimensions in this thesis

The present research is concerned with developing computerised learning environments that employ advanced speech technology to facilitate L2 learning. Therefore, we are primarily concerned with the affective experience of learning that takes place in such environments. Ushioda's research obtained information about learner experiences by means of the stories, reflections or explanations students provided through open questionnaires or interviews. From these, it was

possible to explore the qualities of learning experiences (or other L2-related experiences) that are perceived by learners to have a positive motivational impact. For these reasons, and its attention to time and the role of observations, Ushioda's work has had a strong influence on the perspective we take to research affective dimensions of CALL-based learning.

At the same time, our aim is to work with data that is readily processable with computer technology. This is because, in a CALL context, computer programs could potentially use affective information from learners to adapt characteristics of their practice environment. In this respect, our approach has similarities to work by MacIntyre and Legatto (2011) who have developed a quantitative computer-based self-report method to track changing levels of affective state. In an L2 speaking context, they have shown that willingness to communicate, anxiety and approach-avoidance motivation fluctuate during L2 use and in relation to task difficulty. In this thesis, we focus on qualities of affective (emotion-related) learning experiences that learners seem to find motivating, i.e. when students enjoy practice, feel energetic and competent (Ushioda, 1993).

Finally, although this thesis will focus on affective dimensions thought to be sensitive to the language learning environment, past work has repeatedly shown that more stable affective characteristics can be strong predictors of intended effort or motivated behaviour (e.g. Csizér & Lukács, 2010; Kim, 2012). Therefore, we expect that more global attributes, such as integrativeness or ideal and ought-to L2 self-guides, can influence a learner's experience of and behaviour in their learning environment. In this thesis we explore the long-term motivational characteristics of people who complete spoken L2 practice with our system, from the perspective of Dörnyei's L2MSS (Dörnyei, 2009a).

1.2 Computer-assisted language learning

Innovations in technology allow new kinds of language learning environments. They improve or enrich the kinds of interactions that are possible with computer programs. For example, mature computer technologies like skype, blogs, twitter or video chats can be integrated into a learning context. Another example are specially-built computer-mediated communication or computer-supported collaborative learning environments (Volle, 2005) that provide new spaces for learners and instructors to interact with each other. Another branch, called tutorial CALL (Heift & Schulze, 2015), tries to simulate the features of practice that make one-on-one human tutoring so effective (Bloom, 1984) and make them available to students in self-study. Some systems support natural language interactions (Eskenazi, 2009), or employ technology that can automatically provide corrective feedback (CF) and adapt to students' strengths and weaknesses (Heift & Schulze, 2007), while others can adapt to students' interests (Heilman, Collins-Thompson, Eskenazi, Juffs, & Wilson, 2010). The rationale for this is that it is not always feasible, convenient or

affordable for people to find human tutors. Overall, with technology and the internet in much of the world's homes, people learning languages have access to a variety of computer-based environments for L2 learning.

For researchers, strong arguments have been made for computer environments as important tools for scientific inquiry into language learning, i.e. for understanding pedagogical effectiveness and learning processes (Chapelle, 2001; Presson, Davy, & MacWhinney, 2013). Chapelle (2001) outlines two types of research that can be useful: research into learning conditions and research into language assessment. It is the first type, where the aim is to “gather evidence about the effects of instructional conditions of learning” (Chapelle, 2001, p. 132), that is most relevant here. At the core of this type of research are questions related to pedagogical effectiveness: Does a given study environment, teaching approach or practice feature impart learning gains? By what process are the learning gains realised? How do the learning gains of one environment, technique or practice feature compare to those of another?

Using CALL technologies, researchers can investigate these questions by designing activities that operationalise learning conditions precisely. The activities can be embedded in a computer environment, which provides a high degree of control over what actions take place and when, and makes it possible to automatically collect a variety of data on learners (Presson et al., 2013), including behaviour but also their affective experiences (e.g. Ainley, Hillman, & Hidi, 2002; Moos & Azevedo, 2008). Furthermore, because technology is becoming increasingly adopted in language classrooms, computerised environments are transitioning from being seen as a lab context to an authentic practice environment (Bax, 2003; Chambers & Bax, 2006).

While there are many issues for CALL research that are important and deserve more attention, in our view, there are two areas which are promising yet have received relatively less attention in the literature: the first area is to increase our understanding of how the development of L2 spoken proficiency can be facilitated in spoken computer-based language practice, while a second important area has to do with exploiting functionalities in CALL to understand how affective dimensions are influenced by learning conditions over time. We elaborate on each of these below.

1.2.1 CALL research targeting the development of spoken L2 proficiency

Speaking can be one of the most sought-after abilities when it comes to L2 learning (Bowden, Starrs, & Quinn, 1989; Busse & Walter, 2013). At the same time, it is also a difficult skill to acquire, due to the additional demands of articulation and temporal social-interaction constraints (Bygate, 1987). Providing learners with adequate practice opportunities in the classroom is complicated by difficulties in realising ideal conditions for speaking practice (e.g. individualised corrective feedback) due to practical constraints. Speaking in the L2 classroom can also provoke feelings of anxiety which may inhibit

learners from participating and the effectiveness of practice (Horwitz, Horwitz, & Cope, 1986; MacIntyre & Gregersen, 2012; Öztürk & Gürbüz, 2014; Sheen, 2008). Further, while some research suggests that skills acquired in the written modality are transferable to the spoken modality (Payne & Whitney, 2002), application of skill acquisition theory (DeKeyser, 2007) would suggest that many skills are not transferable, with the implication that the most effective way to learn to speak an L2 is to practice speaking.

Computerised learning environments, which are often associated with providing a comfortable learning context (e.g. Warschauer, 1996), can support L2 speaking practice in different ways. Environments that support Computer-mediated Communication (CMC) can be used for speaking practice (Jauregi, de Graaff, van den Bergh, & Kriz, 2012), through synchronous or asynchronous voice or video chats (Gleason & Suvorov, 2012). Another possibility, sometimes considered “the Holy Grail of CALL” (Bush, 2008, p. 461), is Automatic Speech Recognition (ASR), as it has the potential to enable L2 speaking self-study. ASR is the technology developed to allow spoken human-computer interactions. Speaking generally, ASR accepts a spoken utterance and uses signal processing techniques coupled with statistical models of different elements of language to search for a sequence of written words that best corresponds to the utterance. Once a spoken utterance has been translated to a sequence of written words, it becomes a sequence of discrete symbols which can be more readily handled by programming languages. CALL applications can use ASR to provide language learners with the opportunity to speak in one-to-one activities with the computer.

There have been many CALL applications built with ASR, supporting a fairly wide range of L2 speaking skills, such as pronunciation (Neri, 2007), grammar (Penning de Vries, Cucchiari, Bodnar, Strik, & van Hout, 2014b), and pragmatics (Chiu, Liou, & Yeh, 2007), with a similarly wide-range of instructional approaches. ASR has been used in combination with advances in hardware performance and other software technologies to create virtual worlds where learners can have conversations with virtual L2 speakers (Johnson & Valente, 2009; Morton & Jack, 2010). ASR has also made it possible to get corrective feedback on the speech a learner produces (e.g. Neri, 2007; Penning de Vries, 2015, see also Chapter 3).

Despite these accomplishments, there are some views in the literature that are sceptical about the appropriateness of using ASR technology with L2 learners (Coniam, 1998, 1999; Derwing, Munro, & Carbonaro, 2000) or cautious about ASR-enabled CALL systems (Clifford & Granoien, 2008). Certainly, some uses of ASR for providing detailed feedback have not been successful, because the intended functionality could not be provided by ASR with the necessary accuracy (e.g. Herron & Menzel, 2000). However, this does not mean that we should dismiss the ability of ASR to contribute meaningfully to language learning. As Bush (2008) writes:

Such an inference, however, is exactly the wrong lesson to draw from Clifford and Granoien’s insightful overview of the serious

limitations of much of the current generation of CALL. Although they recognize the limitations of today's CALL in general and speech recognition technology in particular, we should not conclude that because CALL cannot do everything, it cannot do anything useful in contributing to the general language acquisition process.

Over the last three decades, a better understanding has been gained of the limitations and affordances of ASR (Strik, Neri, & Cucchiari, 2008). The chances for having ASR that performs well can be addressed with different strategies for applications that use ASR. For example, the statistical models can be trained on a corpus of speech that is representative of the type of speech the application will be processing.

A second way is by designing the activities to be within the limits of the ASR (e.g. Strik, Cornillie, Colpaert, Doremalen, & Cucchiari, 2009). One factor that can determine the feasibility of including speech-interactivity is the type of language activity that one aims to develop. Speaking generally, in activities with "open response designs" (Ehsani & Knodt, 1998, p. 55) where the learner expects to be able to respond naturally, such as role-plays or virtual conversations, it can be relatively difficult for ASR to achieve word-for-word accuracy. However, in communicative activities such as these, this level of accuracy may also be unnecessary (Ehsani & Knodt, 1998). Rather, what is required is that the user's utterance be correctly understood pragmatically, so that the system responds in an appropriate way (Sagae, Johnson, & Bodnar, 2010). Further, as Cook (1988) points out, beginning learners can have a comparably limited range of language structures available. From a pedagogical view, creating activities which are close-ended (by, for example, requiring that learners choose their utterance from a list of on-screen alternatives) is often acceptable and even pedagogically advisable, and at the same time simplifies the recognition task for the ASR engine. In sum, developers interested in creating systems that provide CF on oral L2 productions must carefully consider the limitations of the technology and find a balance between those limitations and the planned language learning activities.

Therefore, while it is true that speech technology is "not yet sufficiently reliable to allow for free conversations" (Heift & Chapelle, 2012, p. 558), we argue that ASR can offer beneficial practice for other spoken language skills. A more nuanced view of spoken language practice and fine-grained look at speech technology can reveal that, even with its limitations, speech and dialogue technologies can be used to create pedagogically effective learning environments.

When sufficient attention is given to its limitations, the combination of ASR with a CALL environment provides a compelling method for researching topics related to the development of spoken L2 proficiency. The bulk of studies investigating learning effectiveness and processes have conducted research on practice in the written language modality (e.g. Heift & Chapelle, 2012; Heift & Schulze, 2007), while explorations of the effectiveness and learning processes in CALL speaking practice are comparably rare (Felix, 2008). Reviews with a

focus on speech-interactive environments have tended to highlight the spoken interactions that can be supported and the performance of the technology (e.g. Ehsani & Knodt, 1998; Gamper & Knapp, 2002; Wachowicz & Scott, 1999). Holland and Fisher (2007) is an important resource for situating speech technology into different phases of CALL research, with the enlightening discussion of speech technology from the point of view of second language acquisition mentioned above (Clifford & Granoien, 2008). A few studies have investigated the effectiveness of speech-interactive systems (e.g. Cucchiarini, Neri, & Strik, 2009; Neri, 2007; Strik, van Doremalen, Colpaert, & Cucchiarini, 2013). However, in general, there is a need for additional research that investigates spoken L2 practice.

1.2.2 CALL research with attention to affective dimensions

Affective dimensions are an important dimension in CALL research and evaluations. This is made clear by a recent meta-analysis of CALL effectiveness research targeting reading, writing and spelling from 1981 - 2005 (Felix, 2008). The study provides a comprehensive overview of how the field has approached evaluating CALL effectiveness. Relevant here are the findings concerning learners' perceptions of CALL which observe that research into "student perceptions of CALL represents the largest area of activity to date" (p. 147) with analyses indicating that "use of technologies can have a positive effect on student attitudes and participation (although not reflected in higher achievement)" (p. 154).

The role of affective dimensions in CALL evaluations is not limited to learners' retrospective perceptions of practice in computerised environments. Instead, researchers have proposed that, for CALL systems to be effective, they should promote certain affective characteristics in learners. For example, Chapelle (2001) outlines "theory concerning ideal cognitive and socio-affective conditions for instructed SLA" (p. 45) to establish a set of principles for evaluating CALL systems. Here, Chapelle suggests that CALL should aim to promote "learner's interest in seeking out opportunities for communication and their willingness to communicate in these situations" (p. 51) and promote the affective conditions that underlie and contribute to WTC. The meta-analyses by Felix (2008) suggests that CALL environments can have this effect, finding "some indication that CALL had positive effects on motivation; computer literacy; target culture awareness; reading and listening skills; classroom climate; comfort; and participation" (p. 155).

Work by Izquierdo, Simard, and Pulido (2015) provides a recent example of a study investigating the potential impact of CALL on learners' affective characteristics which at the same time shows that affective and motivational impacts can be different for learners with different profiles. The study builds on Gardner's (1985b) socio-educational model to investigate how learner attitudes towards L2 learning and towards their L2 class are impacted by two different

independent variables, namely learner proficiency and two types of multimedia instruction: a meaning-based condition in which the target linguistic feature is not made explicit and another where the feature is made explicit. Analysis of the pre- and post-instruction questionnaire data reveal that the attitudes of low-proficiency learners remained stable. For medium proficiency learners, the study finds that learners who received the meaning-based instruction report more positive attitudes post-instruction, while attitudes remained stable for the group receiving instruction in which the pedagogical focus was explicit.

In a recent CALL conference in 2010 titled ‘Motivation and beyond’, affective dimensions of language learning were in the spotlight (Aerts, Colpaert, & Oberhofer, 2010). The proceedings had theoretical papers, as well as studies on motivation and affect, and involved a variety of methodologies and CALL learning situations, including CMC environments (D. N. Brown, 2010), distance learning (Caron & Debeurme, 2010), classrooms (Dembovskaia, 2010), gaming (Cornillie, Vandewaetere, & Jordaens, 2010), and role-play activities in virtual worlds (Ferreira da Silva, García, Edwards, & Domínguez-Noriega, 2010). Many of these conference papers have been further developed and published as CALL journal articles (e.g. Colpaert, 2010; Cornillie, Clarebout, & Desmet, 2012; Gleason & Suvorov, 2012; Jauregi, de Graaff, van den Bergh, & Kriz, 2012; Ushioda, 2011). One observation is that, with the exception of the study by Cornillie et al. (2010), it seems that there were few studies of motivation in self-study environments such as games or tutorial CALL systems, and no such studies with ASR-enabled CALL. Furthermore, both Alm (2010) and Raby (2010) point out that the motivational impact of CALL is an important topic that is in need of further attention.

A point worth mentioning here has to do with the use of self-report questionnaires which are often used in studies that investigate students’ perceptions of CALL practice (for an account of a recently developed questionnaire with psychometric evaluation, see Vandewaetere & Desmet, 2009). Recently, Fischer (2007) discusses the disadvantages of relying solely on students’ perceptions about CALL practice when evaluating CALL effectiveness. A correlational analysis of students’ perceptions of the perceived instructional value of different program components and the self-reported and actual use of said components, captured through logs, revealed very few significant relationships. The study finds that the results “call into serious question the use of student questionnaires and surveys to estimate the effectiveness of CALL software” (p. 429).

While the questionable reliability of self-report questionnaires has been convincingly shown above for learner perceptions concerning their use and valuation of specific CALL application features, it seems worth mentioning that in the case of certain other affective dimensions such as affective responses to the learning environment (e.g. enjoyment, boredom), feelings of motivation, or confidence or self-efficacy, it is exactly learners’ subjective experience that we would like to record and that, for these types of affective dimensions, the most direct way of obtaining that information is by asking the learner (although

advances in affective technologies may provide more objective methods for measuring some of these characteristics).

Interestingly, the limitations of self-reporting and memory have also been discussed by leading L2 motivation researchers (MacIntyre & Gregersen, 2012; Ushioda, 2012). For example, MacIntyre and Gregersen (2012) presents research on affective experiences which has revealed that people’s more global impressions or affective memories are not the simple accumulation of affective experiences, but instead that timing and strength of the affective responses to events can disproportionately impact the general affective character of an experience. Further, Ushioda (2012) notes that researchers using self-reports are limited to what the learner can meaningfully recall. Two important suggestions are made: to attempt to try to capture motivation *in situ*, i.e. at the time of practice, and to complement self-report data with observational data documenting interactions that take place during learning, mirroring the recommendations by Fischer (2007) above.

Together, these ideas suggest that it could be “important to investigate what students do” (Fischer, 2007, p. 427) in relation to what they think or feel. In other words, we may gain potential insights by investigating affective dimensions in CALL using a situated approach that examines learner affective characteristics while the learner is engaged in L2 practice and complements this subjective self-report information with objective logs of practice behaviour. To our knowledge, such a learning-process approach to affective dimensions of language learning has rarely been conducted in a CALL context (see Chapter 2).

1.3 GREET: A platform for researching affective dimensions in spoken CALL practice

This thesis investigates the two areas discussed above, namely affective dimensions of learning to speak an L2 in the context of an ASR-enabled CALL system that provides oral L2 grammar practice. As a step towards studying affective dimensions in a CALL environment, we developed a computer-based experimental platform to support a situated look at affect. The platform, called GREET¹, consists of a variety of activities. Key activities in the system are 1) a computersied L2 practice module, providing the instructional condition, 2) two self-report components designed to elicit affect and motivational profiles from learners, 3) computer logs of learner behaviour and 4) authoring tools. A brief description of each of the activity modules is provided below.

1.3.1 Computerised spoken L2 practice for word order

The practice context we employ provides L2 learners of Dutch with autonomous oral grammar practice. It focuses on one specific aspect of Dutch grammar that

¹Although written using all capitals, the name “GREET” is not an acronym.

is problematic for learners, namely Dutch word order. The exercises are made up of pairs of video clips and question-and-answer activities. Learners first watch part of a story in the video clip. Later, they answer questions about the video by speaking their answers aloud.

Recent research into the effect of corrective feedback has shown that it is beneficial for developing L2 proficiency (Lyster, Saito, & Sato, 2013; Sheen & Lyster, 2010). Furthermore, students typically report wanting to receive corrective feedback (Cathcart & Olsen, 1976; Chenoweth, Day, Chun, & Luppescu, 1983; Jean & Simard, 2011; Plonsky & Mills, 2006), but teachers can be reluctant to provide it (Vásquez & Harvey, 2010). Even if willing, teachers must often deal with classroom constraints that prevent them from providing individualised and well-timed feedback.

In answer to this, a number of CALL researchers have developed feedback modules which enable learners to receive feedback under more ideal conditions (for an overview of feedback in written interactions, see Heift & Schulze, 2007). Providing these conditions in autonomous oral L2 practice requires the use of automatic speech recognition technology (ASR) to convert oral productions to text so they can be automatically evaluated for grammatical correctness. The use of ASR technology in CALL applications is challenging, as a balance between the limitations of the technology and usefulness of the pedagogy must be found (Neri, Cucchiaroni, & Strik, 2003).

To develop learners' word order accuracy while keeping the speech processing task within the limits of the ASR technology, we designed the oral quiz component of practice as a block ordering exercise. Learners first read a question and then attempt to produce a grammatically correct utterance from the word blocks shown to them on the screen. Use of the word blocks improves ASR accuracy by making the learner's oral production more predictable, but still challenges them to practice Dutch word order.

The feedback we have designed is modelled after elicitation feedback (Lyster & Ranta, 1997). If the answer is correct the learners receive confirmatory feedback and in cases when their answer is wrong they are notified and prompted to try again. From their second mistake onwards they also receive a hint from the computer: the next word block is automatically placed in the answer area.

1.3.2 Innovative features enabling learning conditions research from an affective perspective

In-practice self-report for tracking affective states during practice

Because we were interested in motivation during the learning process, we designed the system to monitor how long learners practice and periodically present learners with a short questionnaire for tracking their affective state. Learners typically complete a questionnaire once, before practice begins, and complete additional questionnaires at regular intervals. The interval between

questionnaires is configurable, but in experiments we have scheduled these questionnaires every 15 minutes. When the practice time limit is reached, the system automatically presents the learner with the questionnaire as a web page. After completion, the learner is returned to where they were in the practice material.

The questionnaire consists of three semantic differential scales designed to survey *enjoyment*, learners' affective responses to practice with the system, *willingness-to-practice*, learners' desire to continue practice with the system or stop practice and do something else, and *self-efficacy*, as the learner-estimated perception of their ability to do well in future practice. We chose to use quantitative scales because we wanted to be able to use statistical analysis to measure effects and search for relationships between motivation and behaviour recorded in the log data. A second reason was to have the data computer-processable, to provide a basis for developing modules that automatically analyse learner motivational state and adapt computer behaviour in the future.

Elicitation of Ideal L2 self profiles using a computer-based card sorting activity

Another type of data important to SLA research which could be elicited in a CALL environment are a learner's motives for learning an L2. Though knowledge of such motives can provide valuable context for the design and evaluation of a practice environment (Morton & Jack, 2010), and could serve as a resource for personalising instruction, we are unaware of any attempt to elicit such data using a CALL environment. In response to this gap, the system implements a CALL activity that elicits learners' motives for learning an L2. The basis for the activity is the L2MSS (Dörnyei, 2009a), which views L2 motives in terms of a learner's possible selves embodying their aspirations, goals, and worries with regard to learning an L2. A series of selection and sorting exercises with a deck of electronic possible selves cards elicits learners' hoped-for and feared L2 selves and collects ratings on their desirableness, efficacy and likelihood. The activity is web-based and takes approximately 30 minutes to complete.

1.3.3 Computer practice logs

One of the often-cited advantages of SLA research in computerised environments is the detail with which observations about the learner's behaviour can be logged (Penning de Vries, 2015; Presson et al., 2013). The GREET system is implemented as a web application and has a client-server architecture. Learners interact with the client web page interface, by watching videos, recording utterances and viewing feedback. As they interact, their actions are sent to a dedicated logging server which waits for incoming events and saves them to a database. As the server processes the learner actions and prepares a response, it also logs the results of its own actions, such as the output of the

ASR module or the kind of feedback presented to the user, and records these events as well. To contextualise the event collection, information about which learner was practising and what task they were practising is also saved with the events. All events are also recorded with timestamps, which allow us to look at user behaviour across time. Lastly, the spoken utterances recorded by learners during practice are saved to the server, where they can be annotated with an integrated web tool, or retrieved for linguistic study or ASR training.

1.3.4 Authoring tools for researchers

In their chapter on future directions, Chapelle (2001) offers a number of suggestions for authoring tool functions that would benefit computer-based SLA research. We developed some of these features in the GREET system, to allow us to use it as an experimental platform. A back-end administrator facility was developed to allow researchers to define different experiments. We have developed a courseware framework consisting of tasks and task lists. Tasks in the system follow the usual intuitive definition, i.e. units of work that learners complete, most often to develop their L2 proficiency, but also to provide information to the researchers. These components are modular and can be assembled into task lists (i.e. task sequences) which can then be assigned to the learner to provide them with activities to complete once they log in to the system. Task lists allow researchers to interleave tasks of different types, providing a method of investigating task-related motivational variation (Julkunen, 2001) by obtaining motivational data before and after a task. It also means we can go even more fine-grained by configuring the system to periodically pause practice, transition the learner to a motivation instrument where they complete a brief questionnaire, and then return the learner to practice. Furthermore, progression from task to task is automatic and includes customisable briefing and debriefing pages to smooth the transitions.

A second important feature of the administration facility is that it allows us to have identical practice environments, but change key elements of the practice that are related to our research questions. In our work on the effect of corrective feedback (see Chapter 3 and 4), for example, we have configured different task lists which are different only in that one list contains practice modules that provide corrective feedback, while the other contains the same practice modules with feedback disabled.

1.4 Goals and outline of this thesis

The present study aims to investigate factors influencing learners' affective practice experience in the context of the computerised learning environment we developed. It also aims to study the relationship between the practice experience and relevant objective measures such as initial L2 proficiency and learning outcomes. These efforts form the contents of this dissertation, which is organised into the remaining five chapters.

This is by no means the first study to focus on the subjective practice experience in a CALL environment. Chapter 2 surveys the literature to show the opposite, that this is a topic of interest to the CALL community and that a significant amount of work has been conducted in the field. In surveying the literature, attention is paid to the theoretical frameworks that are used, the methodologies, the treatment of time, the use of computer logs and the use of macro perspectives. We find that there are gaps between how motivation and related affective characteristics are evaluated and the latest motivational theories. An attempt is made to provide useful directions for future research.

In Chapter 3, we present our first study on affective dimensions of learning using the GREET platform. The study examined how one controversial practice characteristic, corrective feedback, impacted three aspects of the affective practice experience: enjoyment, willingness to practice, and self-efficacy, i.e. learners' judgment of their ability to do well in the exercises. The aim is to investigate the relationship between learners' experience and the objective logs of the events that took place during practice, i.e. the student-computer interactions. The study looks closely at the trajectories of learners' enjoyment, willingness and self-efficacy within a practice session and across sessions.

In Chapter 4, we present our second study using GREET, which adopted a more global perspective relative to the previous study to investigate whether automatic corrective feedback provided on learners' word order errors would enhance their practice experience. Additional factors that were examined include proficiency in the L2 before the experiment and the learning outcomes of practising with the system.

Chapter 5 takes on a different perspective. In this chapter, we address other points that emerged from the literature review in Chapter 2 which suggested that computer methods for eliciting motivational data and researching learners' L2 motivation from the perspective of the L2MSS (Dörnyei, 2005, 2009a) are innovative areas of research. This chapter describes how we developed a computer application for eliciting motivational data. The activity is implemented as a card sorting activity for eliciting data on the hopes a learner has for their future with respect to speaking Dutch. The chapter begins with theoretical background. Next a description of the application is given. Finally, the chapter demonstrates use of the application by eliciting and analysing possible selves data obtained from participants in previous experiments. We analyse the data using an inverted factor analysis technique associated with Q methodology with the aim of investigating the diversity of motivational perspectives in our population with application to personalising instruction.

In Chapter 6, the final chapter, we conclude the study with reflections on the work completed in the previous chapters and comments on its limitations. Based on these conclusions, we then offer suggestions for future research.

CHAPTER 2

Evaluating the motivational impact of CALL systems: current practices and future directions

This chapter has been adapted from

Bodnar, S., Cucchiarini, C., Strik, H., & van Hout, R.W.N.M. (2016) Evaluating the motivational impact of CALL systems: current practices and future directions, *Computer Assisted Language Learning*, 29:1, 186-212, doi: 10.1080/09588221.2014.927365

Acknowledgements

We would like to thank our colleague Bart Penning de Vries for his feedback on an earlier version of this manuscript. In addition, we thank two anonymous reviewers whose helpful comments have improved the quality of this manuscript.

Abstract

A major aim of computer-assisted language learning (CALL) is to create computer environments that facilitate students' second language (L2) acquisition. To achieve this aim, CALL employs technological innovations to create novel types of language practice. Evaluations of the new practice types serve the important role of distinguishing effective practice environments from less effective environments, while simultaneously informing educational practices and second language acquisition (SLA) theory. Accordingly, evaluations of CALL systems necessarily deal with multiple criteria. Most researchers would probably agree that motivation is an important criterion in CALL evaluations: a system can provide sufficient L2 input and opportunities for L2 output, yet fail to be pedagogically effective if learners are unwilling to participate. Furthermore, knowledge of the motivational impact of practice can provide valuable context linking individual language learners, practice effort and learning outcomes. From the perspective of recent theoretical developments in L2 motivation theory, this paper surveys a representative sample of CALL system evaluations that include motivational impact. Our analysis suggests not only that CALL needs to do more to align its treatment of motivation with recent L2 motivation theories, but also that it is well positioned to do so. We find that (1) few CALL studies treat motivation as it relates to practice as a dynamic variable, (2) behavioural practice logs are underexploited and (3) very few evaluations take into account learners' individual interests and goals. Drawing on these and other findings, we suggest four new directions for developing the motivation dimension in CALL evaluations.

Keywords: CALL, evaluation, motivation, review

2.1 Introduction

Advances in computer technology have resulted in applications that provide innovative practice environments to second language (L2) learners. Developments in automatic speech recognition, natural language processing and student modelling have contributed features to spoken and written language practice in intelligent computer-assisted language learning (ICALL) programs (Gamper & Knapp, 2002), such as corrective feedback (Cucchiarini, Neri, & Strik, 2009; Heift & Schulze, 2007), meaning-focused practice (Johnson, 2010; Morton & Jack, 2010) and individualised instruction (Heift, 2008; Heilman, Collins-Thompson, Eskenazi, Juffs, & Wilson, 2010). Complementing these developments has been the popularisation of technologies for communication over the internet, using video conferencing tools (Develotte, Guichona, & Vincenta, 2010; Jauregi, de Graaff, van den Bergh, & Kriz, 2012), internet forums (Qian & McCormick, 2014; Thang & Bidmeshki, 2010), chat (Sauro, 2012) or web publishing (J. C. Chen & Brown, 2011), providing new possibilities for meaningful computer-mediated communication (CMC) between learners and native and non-native speakers. In this paper, we will refer to both of these types of computer programs offering language practice with the general term *CALL systems*. Though much practice with CALL systems takes place as blended learning (i.e. learning that includes both a classroom or face-to-face instructional component and a CALL system component, Motteram & Sharma, 2009), this paper focuses on the CALL system in such learning environments. Evaluations of new CALL systems serve the important role of distinguishing between effective and less-effective practice while simultaneously informing educational practices and SLA theory (for a recent review of the effectiveness of practice with different CALL technologies, see Golonka, Bowles, Frank, Richards, & Freynik, 2014). Accordingly, evaluations of CALL systems necessarily deal with multiple criteria. They must deal with issues related to learning, such as measuring the impact of practice on a learner's progress towards L2 acquisition, and with issues related to technological performance, such as the accuracy of automatic speech recognition components in processing L2 input (Cucchiarini, Doremalen, & Strik, 2012; Morton & Jack, 2010).

Another important issue in CALL system evaluations is related to learner motivation, which can be defined as their “state of cognitive and emotional arousal, which leads to a conscious decision to act, and which gives rise to a period of sustained intellectual and/or physical effort in order to attain a previously set goal (or goals)” (Williams & Burden, 1997, p. 120). It is the complex and multi-faceted concept that is “responsible for why people decide to do something, how long they are willing to sustain the activity, (and) how hard they are going to pursue it” (Dörnyei & Ushioda, 2011, p. 4). It seems straightforward to reason that a learner's motivational experience during practice has an influence on their actual practice behaviour, and eventual learning gains. To understand the effects of practice, evaluations of CALL systems should include analyses of students' motivation.

Recently, there have been developments in L2 motivation theory which have contributed new perspectives particularly applicable to CALL (for discussions of developments in L2 motivation theory over the past 50 years, see Dörnyei & Ushioda, 2011; Ellis, 1994; Williams & Burden, 1997). For a long period, the study of motivation in an L2 context was dominated by a social-psychological approach that searched for stable and general personal characteristics that predict L2 achievement. Since the early 1990s, however, L2 motivation research has been influenced by educational and motivational psychology and stressed more situated aspects of motivation. Recent perspectives emphasise that (1) the learner's motivational state is best seen as a dynamic variable fluctuating during practice; (2) motivation should be examined in the context of events which occur during practice; and (3) the learner's practice experience is influenced by a complex combination of their present self-concept and their aspirations for the future (Dörnyei, 2009a; Ushioda, 2009b; Ushioda & Dörnyei, 2012).

A growing number of studies in the CALL literature are including motivational aspects in their evaluations of CALL systems. However, there is a gap between what recent theories recommend and current practices in evaluations of motivational impact. In this paper, we survey a selection of these publications with the aim of assessing current approaches to evaluating motivation in a CALL setting, using criteria derived from the aforementioned recent developments in L2 motivation theory. In the next section, we begin by reviewing developments in L2 motivation theory over the past 50 years, spending more time on two views of motivation that have emerged in the last decade. Following this, we present the results of our survey, with attention to the theories used to conceptualise motivation and the methods used to collect motivational data. Based on the results of the survey, we then propose four directions for future research, before presenting our final conclusion.

2.2 Perspectives on L2 motivation

In preparation for our review of studies that evaluate the motivational impact of CALL systems, this section provides background on theories of L2 motivation. We provide an overview of theories employed in CALL research, and also include a selection of the most recent theories not found in the evaluations which, in our view, suggest compelling future directions for improving CALL evaluations.

2.2.1 Early social-psychological perspectives

Initial work on L2 motivation began in the late 1950s and approached researching motivation from the perspective of social psychology. Research of this type was primarily concerned with finding relatively permanent affective individual difference variables that could predict SLA achievement (Ushioda, 2009a). Examples of this perspective include Gardner's socio-educational model

of SLA (Gardner, 1985b), highlighting the role of attitudes towards native speaker populations, and work by Clément (e.g. Clément, 1986), investigating the role of linguistic self-confidence. In this paper, we will refer to this kind of motivation¹ as *global motivation*, to emphasise its relative permanence across different practice situations.

Gardner and associates studied the motivation of foreign and second language learners in the social context of Canada. Their work is perhaps best known for examining the role of integrative and instrumental orientations in SLA achievement. An *integrative orientation* can be summarised as a sincere desire to be like or become close with the community of target language speakers, while an *instrumental orientation* refers to more pragmatic benefits of learning a foreign language, such as getting a better job. Initial studies by Gardner indicated that learners who were integratively oriented tended to be more successful at learning a foreign language than those whose motives were instrumental.

A major contribution of Gardner's work is the Attitude/Motivation Test Battery (AMTB), a comprehensive questionnaire instrument designed to obtain self-report data on a variety of motivation-related affective constructs. This instrument has had a very significant role in L2 motivation studies in Gardner's own work, and continues to serve as a starting point for studies researching motivation from more recent perspectives such as the L2 Motivational Self System (e.g. MacIntyre, Mackinnon, & Clément, 2009).

Although social-psychological perspectives have made major contributions and have been regarded as very advanced for their time for their acknowledgement of the role of social context, over the years, there has also been much discussion of their shortcomings (Dörnyei, 2005; Dörnyei & Ushioda, 2009). Because of its relative dominance of the field, Gardner's Socio-Educational model has received much criticism (Ellis, 1994).

One particularly relevant criticism of Gardner's model is that the power of integrative motivation as a predictor of SLA achievement did not always generalise to other learning contexts. In the French-as-a-second-language context of Canada, where Gardner conducted the initial research that led to his theory, attitudes towards a particular native-speaker population (French Canadians) had a large role in predicting SLA achievement. However, much research exists which finds contexts in which orientations other than integrative ones, such as instrumental orientations (Ellis, 1994), and even other affective variables such as self-confidence (Clément, 1986), can be better predictors of SLA achievement, or that SLA achievement can occur without positive relationships with integrative orientation, or even with negative ones (Au, 1988).

¹Previous literature has organised self-esteem (H. Brown, 1981) and later motivation (H. D. Brown, 2000; Ellis, 1985) into global, situational and task levels. In this review we organise the research using the categories of global and situated motivation to align with recent literature distinguishing between macro- and micro-perspectives on L2 motivation (e.g. Kissau, 2006; Ushioda, 2009b; Ushioda & Dörnyei, 2012).

2.2.2 Alternative perspectives on global L2 motivation

As SLA researchers began to note the shortcomings of social-psychological perspectives, they began to explore applications of other frameworks. We provide a brief description of the most influential of these which also play a role in CALL motivation research.

Social cognitive learning theory (Bandura, 1986) centres on an individual's self-beliefs and self-knowledge. Two important concepts from the theory that have played a role in the research are self-regulation and self-efficacy. *Self-regulation* describes a person's "mechanisms for perception, evaluation and regulation of behaviour" (Bandura, 1978, p. 348). Reflection on past behaviour translates into the regulation of future behaviour, such as when, for example, a person decides whether or not to continue learning a second language. Influencing the regulation of behaviour is *self-efficacy*, defined as the beliefs one has about the degree to which one is competent and capable. Self-efficacy beliefs can impact motivation by influencing a learner's expectations regarding the outcome of an action: stronger self-efficacy beliefs may make a successful outcome seem achievable and thereby increase the likelihood of a learner engaging in an activity.

Self-determination theory (SDT) (Deci & Ryan, 1985) focuses on learner-internal motivational states described by a continuum of motivational types. At one extreme is intrinsic motivation, the state of being motivated simply because an activity is enjoyable or interesting. At the opposite extreme lies the state of amotivation, a state of complete disinterest in an activity. In between are different types of extrinsic motivations, which are each more or less internalised, that is, more or less in line with the beliefs, goals and values of the individual. The motivational state of a learner is related to their perception of a task and a comparison of that task with their interests and goals. As the task becomes more in line with the goals and interests of the learner, the state of motivation becomes increasingly intrinsic, and in principle, results in more exerted effort.

Recently, the work of Dörnyei (2005, 2009a) has attracted attention for addressing shortcomings in the generalisability of socio-psychological approaches with a theoretical framework called the L2 Motivational Self System. Relying on data from a very large and longitudinal study in Hungary (Dörnyei, Csizér, & Németh, 2006), this framework draws on a thread of research in motivational psychology called *possible selves* (Higgins, 1987; Markus & Nurius, 1986)). The model hypothesises the existence of two important future visions, called *self-guides* in the literature: the Ideal L2 self captures the internalised hopes of the individual; the Ought-to self represents the self one feels obligated to become due to societal responsibilities. An awareness of the discrepancies between desired future self-guides and the perceived plausibility of those self-guides, together with a person's current experience of L2 learning, yields a motivational state.

A point to be made about studies with a focus on global L2 motivation in general is that their treatment of motivation clashes with educators views

of the concept (Crookes & Schmidt, 1991). In these approaches, the research concerns are long-term and macro-contextual, and motivation is synonymous with motivational orientations that are assumed to be relatively stable and trait-like, as evidenced by the research designs in which questionnaires were administered once and related to other data sources such as final grades. At the same time, language teachers in classroom settings would probably agree that motivation is more closely related to learners' levels of attention, participation, interest and effort (Crookes & Schmidt, 1991). For educators, it is also probably generally accepted that teachers, classmates and the classroom, in other words, the every-day context in which the language is learned, have an impact on the motivation of the learner. As Ellis (1994) writes, teachers tend to “see it as their job to motivate students by engaging their interest in classroom activities” (p. 516).

2.2.3 Perspectives on situated L2 motivation

In the early 1990s, an influential article by Crookes and Schmidt (1991) made a strong case for shifting the focus of motivation research to issues that are more relevant to everyday classroom education. Research with this focus investigates “the ongoing situated processes shaping day-to-day motivation such as the influence of the teacher or social learning environment” (Ushioda, 2011, p. 201). In this paper, we will use the term *situated motivation* to refer to this type of day-to-day situation-specific view of motivation. Below we highlight theories on situated motivation which play a role in CALL research.

Dörnyei and Ottó's Process Model of L2 motivation (1998) is a synthesis of existing motivation theories into a model sufficiently comprehensive to explain classroom L2 motivation. A major contribution of this theory is the detail with which it models the dynamic change of motivation. Put briefly, the cognitive processes involved in a learner's motivational system occur in one of three temporal phases: The pre-actional phase involves goal setting and intention formation and initiation; the actional phase includes processes which regulate action towards the completion of goals; the post-actional phase involves reflective activities such as attributing outcomes to causes, formation of strategies, and modification of standards. Specific motivational influences relevant to each phase are included in the model.

Sociocultural theory (Vygotsky, 1978, originally 1934) is primarily known as a theory of learning, but it also applies to L2 motivation. As Ushioda (2007) points out, the theory presents a view of motivation that emphasises the role of social interactions. Learners are assumed to have an innate need to “be competent and effective in their interactions with the world” (p. 10). Teachers channel this need by providing interactions within a learner's zone of proximal development. Interactions are carefully managed to address not only cognitive concerns, but also motivational ones (p. 11):

Recruiting interest in the task, maintaining pursuit of the goal, and controlling frustration during problem solving focus very much on

the motivational process of task engagement and seem to constitute what might be called motivational scaffolding.

Over time, success in the scaffolded interactions leads learners to internalise goals presented in tasks and also recognise their own agency and develop their ability to self-regulate.

A recent approach that merits closer consideration is Ushioda's 'person-in-context relational view' of motivation (Ushioda, 2009b) which can be seen to complement Dörnyei's work on global L2 self-guides by focusing on learner motivation in the micro context of language practice. In Ushioda's words, the view "tries to capture how motivation develops through and shapes interactions among persons in a specific context" (Ushioda, 2012, p. 68). Of particular interest to this review are Ushioda's suggestions for empirical methods for studying motivation in these complex learner-context interactions. Typical self-report or interview techniques are retrospective (i.e. temporally distant) and limited to what participants can recall and articulate. Ushioda (2012) suggests complementing these data sources with objective records of learner-context interactions during practice to capture events as they occur. In the author's words "a clear direction for future research must be an empirical focus on the dynamic and interactive contexts through which motivation processes evolve" (p. 70).

A second point to mention reflects Ushioda's position regarding the importance of conducting research relevant to classroom teachers. Ushioda (2009a, 2011, 2012) stresses that researchers will have to abandon the reduction of students to abstract language learners and instead view them as real people in order to generate research findings applicable to an educational setting. They explain how the many identities of a student can be transported into the language classroom to engage students (Ushioda, 2011). By paying attention to personal details, such as hobbies, interests, careers, and histories, their instruction can be personalised to sustain or enhance their motivation.

2.3 From theory to CALL practice: key ideas for evaluating motivational impact

Taken together, the developments in L2 motivation theory described above suggest a number of ideas for researchers evaluating the motivational impact of CALL systems:

1. CALL systems, as practice contexts, have an impact on the situated motivational state of learners. Because knowledge of the motivational impact can help to explain practice behaviours and learning gains, researchers should include situated motivation in evaluations of CALL systems.

2. Studies should adopt a dynamic view of motivation and employ methods that capture changes over time. Straightforward examples of how this can be approached include asking learners to report their levels of motivation or other affective states like confidence or anxiety periodically in interviews (e.g. Busse & Walter, 2013), or when changes occur, using questionnaires (e.g. de Vicente & Pain, 1999) or journals (e.g. Raby, 2007). Another innovative approach is to have learners review video recordings of their language practice and annotate the videos with descriptions of their motivation or anxiety (e.g. MacIntyre & Legatto, 2011).
3. It is important that evaluations complement subjective situated motivational data with records of learner behaviours during practice. In this respect evaluations of CALL systems should employ records of learner actions and system responses to search for links between types of learner-system interaction sequences and motivational changes. Generally speaking, the interaction sequences that we believe will be of interest will be those that suggest different learner states, such as engagement, struggle, frustration, hesitation, confusion or uncertainty, or effort and perseverance, to name a few examples. The individual actions making up these interactions will vary with practice environment and language task. For example, in CALL environments, the interactions of interest will be the button clicks, page views, question attempts and other behaviours performed via the graphical user interface and the system's subsequent responses. In blended learning contexts where learners practice outside of computer environments, human observations of question asking, error correction, participation in group discussions and other learner-learner and learner-teacher interactions could contribute records of learner behaviour indicative of motivational state.
4. Evaluations of CALL systems should include learners' global motivational characteristics, as this information could provide insights into the influence of global motives on learners' situated motivation, practice behaviour and L2 proficiency gains. New research could investigate these global-situated relationships by employing a top-down approach that begins by first profiling learners' global motivational characteristics. In our view, both quantitative and qualitative techniques (such as questionnaires and interviews) could be successfully employed. Next, groups could be formed using the global motivational characteristics found to vary within the surveyed population (such as levels of intrinsic or extrinsic motivation). Finally, the research could search for within-group patterns and between-group differences in situated motivation, practice behaviour, and proficiency outcomes.

In the next section, we examine the extent to which these ideas for evaluating the motivational impact of practice can be found in CALL research.

2.4 Evaluating the motivational impact of CALL systems

Above, we presented four ideas for evaluating the motivational impact of CALL systems derived from recent developments in L2 motivation research. We believe that application of these ideas could provide useful insights into the links between individual language learners, their experience practising with particular CALL systems, and their learning outcomes. This section presents a literature review aimed at comparing the ideas presented above with the approaches taken in published evaluations of CALL systems. An initial keyword search using Google Scholar and journal-provided search engines for different venues² where CALL research is published yielded approximately 80 studies which investigate motivation-related issues in a computerised instructional setting. From these 80, we selected studies which came closest to realising the ideas of recent L2 motivation theories described in the previous section. Studies investigating aspects of L2 learning (as opposed to math or computer programming skills) and meeting one or more of the following criteria were included in the review: (1) they investigate situated motivation, or a related construct such as anxiety or self-efficacy; (2) they treat motivation as a dynamic variable; (3) they record observations during practice; (4) they investigate global motivation as either an independent (i.e. explanatory) or dependent variable. Based on these criteria, we selected 22 studies, published between 1999 and 2012, to use as the basis of this review. Table 2.1 provides an overview of our selections, consisting of 10 studies which focus on situated motivation, seven studies which include both situated and global motivation, and five studies which focus on global motivation.

²Sources included primarily Computer Assisted Language Learning, ReCALL and CALICO Journal, but also Computers Education, International Journal of Artificial Intelligence in Education and User Modeling and User-Adapted Interaction. We also considered a number of articles published in proceedings for conferences such as the International Conference on Artificial Intelligence in Education and the International Conference on Intelligent Tutoring Systems. Finally, during the peer evaluation process, two studies were suggested by the reviewers. Due to their relevance, we have also included these articles, published in *Innovation in Language Learning and Teaching* and *The Canadian Journal of Applied Linguistics*, in this review.

Table 2.1: An overview of the 22 CALL evaluations selected for this review. Approaches (Appr.): Quantitative (Qn), Qualitative (Ql), Mixed (M); Participants (N): Adults (A), Children (C).

Situated perspectives	Appr.	Research objective	N	Theoretical framework(s)	Tools for collecting motivational data
J. C. Chen and Brown (2011)	Ql	Attitudes and perceptions about collaborative online L2 writing with an authentic audience	6 A	Self-determination theory, L2 possible selves	Semi-structured interviews
Colpaert (2010)	Ql	Points of conflict between situated personal goals and pedagogical goals in CALL practice	Various groups	Educational engineering	Focus group discussions
Fidaoui, Bahousa, and Bacha (2010)	M	Perceptions and attitudes towards collaborative L2 writing practice in CALL lab	48 C	Intrinsic and extrinsic motivation	Questionnaire, focus groups, interviews, observations
Lehtonen and Tuomainen (2003)	Ql	Suitability of computer-supported online L2 collaborative writing practice	18 A	Sociocultural theory, intrinsic motivation	Retrospective feedback form, email exchange logs
Raby (2007)	Ql	Motivation in autonomous language learning in a language lab with support from tutors	6 A	Social learning theory, process model	Video recordings, observations, journals

Table 2.1: (*Continued*) An overview of the 22 CALL evaluations selected for this review. Approaches (Appr.): Quantitative (Qn), Qualitative (Ql), Mixed (M); Participants (N): Adults (A), Children (C).

Situated perspectives	Appr.	Research objective	N	Theoretical framework(s)	Tools for collecting motivational data
Thang and Bidmeshki (2010)	M	Perceptions of online L2 reading practice targeting science and technology texts	23 A	Social constructivism, process model	Observations, questionnaire, interviews
Ushioda (2000)	Ql	Motivation in an email language exchange between learners with complementary L1s	34 A	Intrinsic motivation and autonomy	Email exchange logs, open-ended post questionnaire
Vinther (2011)	Ql	Motivation in an email language exchange between learners with differing L1s	15 A	Sociocultural theory, autonomy, authenticity	Email exchange logs
Yang (2011)	M	Learner engagement in autonomous and CMC L2 listening and writing practice	118 A	Engagement	Observations, questionnaire, CMC discussion logs
L. Chang (2007)	Qn	Attitudes and self-efficacy in autonomous oral L2 CALL practice	17 A	Social cognitive learning theory	CALL attitudes and self-efficacy questionnaires

Table 2.1: (*Continued*) An overview of the 22 CALL evaluations selected for this review. Approaches (Appr.): Quantitative (Qn), Qualitative (Ql), Mixed (M); Participants (N): Adults (A), Children (C).

Combined perspectives	Appr.	Research objective	N	Theoretical framework(s)	Tools for collecting motivational data
Heilman et al. (2010)	Qn	Impact of a personalisation strategy on interest in reading for vocabulary practice	44 A	Interest	Personal survey, interest exit
Jauregi et al. (2012)	Qn	Attitudes towards interacting with native speakers, course and L2 culture in oral L2 practice using video-conferencing with native speaker peers	36 A	Mixed (socio-psychological)	Questionnaire
Kissau (2012)	M	Self-efficacy outcomes of online and classroom-based L2 teaching methodology courses	62 A	Social learning theory	Questionnaire, interviews
Choi and Johnson (2005)	Qn	Motivation in autonomous conversation practice in a speech-interactive virtual world	6 A	Socio-educational model	AMTB
Van Aacken (1999)	M	Motivation in autonomous CALL practice for learning Japanese Kanji characters	6 A	Socio-educational model, self-determination theory	Motivation and attitudes towards CALL system questionnaires, interviews

Table 2.1: (*Continued*) An overview of the 22 CALL evaluations selected for this review. Approaches (Appr.): Quantitative (Qn), Qualitative (Ql), Mixed (M); Participants (N): Adults (A), Children (C).

Combined perspectives	Appr.	Research objective	N	Theoretical framework(s)	Tools for collecting motivational data
Tennant and Gardner (2004)	Qn	Motivation in L2 reading comprehension practice	77 A	Socio-educational model	AMTB, motivation and anxiety thermometers
Ushida (2005)	M	Motivation in L2 classroom lessons supplemented with online activities	30 A	Socio-educational model	AMTB, computer attitudes questionnaire, participation, interviews, email diaries

Table 2.1: (*Continued*) An overview of the 22 CALL evaluations selected for this review. Approaches (Appr.): Quantitative (Qn), Qualitative (Ql), Mixed (M); Participants (N): Adults (A), Children (C).

Global perspectives	Appr.	Research objective	N	Theoretical framework(s)	Tools for collecting motivational data
M. Chang (2005)	Qn	Intrinsic goal orientation, perception of task value, control of learning beliefs and self-efficacy beliefs in a college-level course designed to teach learning strategies with an online component	28 A	Social learning theory	Questionnaire
M.-M. (2010)	Chang Qn	Perception of task value and control of learning beliefs in a web-based course on L2 reading and grammar	90 A	Social learning theory	Questionnaire
Wehner, Gump, and Downey (2011)	Qn	Motivation in CMC oral practice with native speakers in a virtual world environment	40 A	Socio-educational model	AMTB
Morton and Jack (2010)	Qn	Motivation for learning an L2 and attitudes towards CALL practice in autonomous conversation practice in a speech-interactive virtual world	76 youths	Self-determination theory	Attitudes questionnaire, motivation questionnaire
Gleason and Snorov (2012)	M	L2 motivational self system and perception of CALL system effectiveness in asynchronous oral CMC L2 practice	10 A	L2 motivational system	self structured interviews

With recent developments in L2 motivation theory in mind (see previous section), we proceed with the survey in five parts. We begin with a review of the theoretical frameworks employed by CALL to evaluate situated motivation. Following this, we provide an overview of the approaches used to gather situated motivational data and continue with a review of the extent to which CALL treats motivation as a dynamic variable. Next, we discuss the use of computer logs for recording learner and system actions. We then conclude with a brief review of global motivation information in the evaluations.

2.4.1 Theoretical frameworks in evaluations of situated motivation

Although the large majority of the evaluations in our selection include situated motivation, few studies employ a theoretical framework that views motivation from a situated perspective, such as the theories presented above in our overview. Studies by Raby (2007) and Thang and Bidmeshki (2010) are based on the Process Model of motivation (Dörnyei & Ottó, 1998). This framework pays particular attention to the time dimension of motivation, assigning cognitive processes to pre-actional, actional and post-actional phases, specifying motivational influences for each phase, and providing a description of the process by which one phase changes to the next. Raby (2007) maps the motivational influence of the tools in the CALL environment to phases in the model. A shortcoming of the work by Thang and Bidmeshki (2010) is that the influence of the Process Model is somewhat shallow. The influence of the framework appears to be limited to motivating the collection of motivational data over time, and the study does not interpret their data in relation to phases in the model.

Studies by Vinther (2011) and Lehtonen and Tuomainen (2003) adopt a sociocultural perspective on motivation. A sociocultural perspective provides a framework for investigating how motivation is affected by social interactions in collaborations in a CMC environment. Both studies examine different interaction patterns recorded from practice and highlight some of the difficulties that can occur with group work, though only the latter study relates the interactions to motivational outcomes, provided, in this case, by self-reports.

A study by Yang (2011) bases an evaluation of their CMC environment on student engagement. Student engagement comes from the field of education research and concerns learners' levels of commitment and involvement in a particular learning activity (Fredricks, Blumenfeld, & Paris, 2004). The construct features dimensions that resemble aspects of recent situated motivation theories, with behavioural engagement capturing student participation and involvement, emotional engagement referring to positive and negative attitudes to the practice environment, and cognitive engagement as the willingness to expend effort towards learning (Fredricks et al., 2004). The study assesses the level of all three types of engagement by analysing students' communications in different practice exercises.

A different approach taken by Colpaert (2010) is to let the theoretical basis emerge, over time, from practical experience. Repeated iterations of an educational engineering process over the past 25 years have identified a concept, called a *personal goal*, which has proved useful for assessing the motivational impact of a CALL system. Although the term may appear to refer to a global motivational construct, personal goals are rather “non-conscious or unconscious volitions related to a specific learning situation” (p. 269); they are linked to “states of mind or feelings” (p. 269) which “spring from attitudes toward the learning situation” (p. 269). The educational engineering process mentioned above employs knowledge of points of conflicts between personal goals, belonging to learners, educators, and others involved, and pedagogical goals, to understand the motivational impact of a system and guide further design. The personal goal concept has emerged from this bottom-up process on its own, without strong ties to existing theoretical frameworks. Thus, it would be interesting to relate this concept to theoretical frameworks for situated motivation, such as the Process Model of motivation (Dörnyei & Ottó, 1998) and the person-in-context relational view (Ushioda, 2009b).

To sum up, we find that only six of the 17 studies that evaluate situated motivation employ situated theories, i.e. theories that provide an account of “how motivation evolves at the actual site of L2 learning” (Ushioda, 2012, p. 68). Which theories do the remaining 11 adopt? The answer is that many use frameworks more associated with global motivation, such as the socio-educational model, SDT and others. In the next section we look at practical approaches to evaluating situated motivation, but we will return to these frameworks again below when we look at the types of global motivation information gathered in CALL.

2.4.2 Approaches to evaluating situated motivation

One way of organising the evaluations of situated motivation is by characterising their overall approach to gathering and analysing data. In this review, we have found that evaluations can be qualitative or quantitative, or that they can combine the two in a mixed-method evaluation.

Qualitative evaluations have an open-ended and exploratory quality and tend to use a definition of motivation that is less formally operationalised. For example, to explore motivation in an email language exchange, Ushioda (2000) searched for aspects of practice that were intrinsically motivating using an open-ended questionnaire. In their sociocultural study, Lehtonen and Tuomainen (2003) used open-ended questionnaires to obtain descriptions of how collaboration with peers affected their motivation in online group work. Work by J. C. Chen and Brown (2011) examined learners’ motivational state along an intrinsic-extrinsic continuum in a web authoring activity by interviewing learners on their experience. Colpaert (2010) describes a focus group technique used in their design process for balancing pedagogical and situated personal goals in a CALL system; the technique elicits aspects of

practice which affect learner motivation, learners' emotional responses to those aspects of practice, and suggestions for improvement. A study by Raby (2007) asked learners to report any factors influencing motivation in study journals, including thoughts regarding learning the language, their practice in an autonomous CALL lab, their teacher, peers, or achievements. Finally, Vinther (2011) examined records of interactions between learners in an email exchange to study motivational impact without obtaining subjective data.

Another approach is to collect and analyse quantitative data. Evaluations of this type specify motivational criteria ahead of time and focus on assessing the impact of a CALL system along those criteria, often using Likert scales or similar methods. Studies by Tennant and Gardner (2004) and Choi and Johnson (2005) measure motivation in this way, with the former recording situated motivational intensity and anxiety, and the latter recording state motivation, mental effort, and situational interest; Heilman et al. (2010) measured interest and perceived difficulty to evaluate the impact of a CALL vocabulary tutor that personalised reading materials, while work by L. Chang (2007) elicited beliefs about the usefulness of a system that included an authentic target-language talk show.

A final possibility is to combine qualitative and quantitative data in a mixed-method study. An advantage of mixed methods is that they often present a comprehensive account of situated motivation, including both a quantitative measure of motivational impact and possible explanations for those outcomes in qualitative reports. In work by Ushida (2005), the combination of quantitative questionnaires and teacher interviews provided information on how different implementations of a language course affected motivation. Fidaoui et al. (2010) evaluated a computer lab for grade school children by having learners complete questionnaires and asking learners and teachers to participate in interviews. Jauregi et al. (2012) included closed and open-ended questions in their questionnaire to evaluate the impact of conversation practice in a video conferencing environment with native speakers. Questionnaire items in this study appear social-psychological and more associated with global motivation; however, the study's use of these to measure situated motivation is justified, as the main feature of practice is the ability to converse with native speakers. Furthermore, the study adopts a dynamic view of motivation (discussed in the next section).

There are a number of other examples of a mixed approach. Van Aacken (1999) surveyed attitudes towards a CALL program for learning Japanese and conducted interviews to obtain descriptions of how participants were using the system. Thang and Bidmeshki (2010) combined interviews with closed and open-ended questionnaire items to measure participants' attitudes to an online course that taught skills and strategies for reading target language scientific texts. Yang (2011) used discourse analysis on logs of student and teacher interactions in combination with a questionnaire to measure participants' level of engagement in CMC activities centred around videos presenting aspects of the target culture. Finally, Kissau (2012) supplemented their quantitative

questionnaire on teaching self-efficacy by interviewing teacher candidates training in online and classroom courses and gained insight into factors affecting their self-efficacy.

Summarising these observations, CALL evaluation of situated motivation appears to be balanced between qualitative explorations and quantitative assessments. Evaluations that focus on situated motivation consistently have a qualitative component yet vary in whether they include a quantitative assessment, while for evaluations that include both situated and global motivation the opposite is true. Thus, the evaluations can be seen as being concerned with both measuring the motivational impact of practice and obtaining information about factors that influence the impact. In the next section we report on how they treat the temporal dimension of motivation.

2.4.3 Situated motivation as a dynamic variable

In CALL, the overall trend in evaluations of situated motivation is towards collecting a single sample after practice has completed (see Table 2.2). Frequently used methods for obtaining a single sample are to administer either quantitative questionnaires (L. Chang, 2007; Fidaoui et al., 2010; Yang, 2011) or qualitative questionnaires (Lehtonen & Tuomainen, 2003; Ushioda, 2000). Another common technique is to use retrospective interviews to assess situated motivation. Post-practice interviews can ask about events that occur during practice (J. C. Chen & Brown, 2011; Fidaoui et al., 2010) or features of practice which impact motivation (Colpaert, 2010; Kissau, 2012; Ushida, 2005).

A shortcoming of single-sample approaches is that they do not collect data which describes in-practice motivational fluctuations. Pointing out a limitation with post-practice interviews, but which, in our view, also applies to post-practice questionnaires, Ushioda (2012) writes “the interview context and interactions are necessarily different and disconnected from the research participant’s L2 learning context and interactions” (p. 68). As Ushioda goes on to point out, post-practice methods are limited to what the learner can recall and may not give us the details of their motivational experience. Thus, we take the risk of missing valuable data regarding motivational changes during practice, especially in the case where behavioural logs are not maintained.

A number of studies address the above issue by taking measurements of the learner’s motivational state at different points in time. By paying attention to when the studies measure motivation, we can separate the studies into two groups. In the first group, studies sample motivation between practice sessions, when the learner has completed practice for the day. Van Aacken (1999) interviewed participants once at the mid-point of the semester-long practice period and again after practice had finished. Thang and Bidmeshki (2010) also interviewed participants twice, once at the end of the course and a second time in a delayed post-practice interview, although the advantage of a delayed interview is debatable for the reasons given above. Other studies sample more frequently, such as the study by Jauregi et al. (2012) which elicited

Table 2.2: Treatment of time in evaluations investigating situated motivation.

Treatment of time	Description	<i>N</i>	Studies
Measured at a single point in time	Questionnaires record a snapshot of L2 motivation; interviews or focus groups encourage learners to reflect on their practice experience	10	L. Chang (2007); Fidaoui et al. (2010); J. C. Chen and Brown (2011); Colpaert (2010); Kissau (2012); Lehtonen and Tuomainen (2003); Ushida (2005); Ushioda (2000); Vinther (2011); Yang (2011)
Between-practice repeated measures	Questionnaires and interviews elicit motivation when learners are not engaged in practice, e.g. after a practice session. Intervals include daily, weekly, or part-way through and at the end of a semester.	4	Choi and Johnson (2005); Jauregi et al. (2012); Thang and Bidmeshki (2010); Van Aacken (1999)
In-practice repeated measures	Questionnaires and journals elicit motivational data from learners while engaged in L2 practice.	3	Heilman et al. (2010); Raby (2007); Tennant and Gardner (2004)

data weekly over a 9-week practice period, and the study by Choi and Johnson (2005), in which learners completed a questionnaire at the end of each practice day in a four-day training period.

In the second group, we find studies which sample motivation while the learner is engaged in practice. Learners participating in the study by Tennant and Gardner (2004) updated their state motivation and anxiety using ‘motometers’ and ‘anxometers’, computer-adapted Likert scale items displayed in the computer program as thermometers. Learners used these tools once at the beginning of practice, but no additional in-practice measures were taken. Another shortcoming here is that these measures of state were taken only twice in a semester-long practice period, at the same time learners completed a miniaturised 13-item version of the AMTB. Heilman et al. (2010) surveyed learners more frequently, asking them to rate their interest level and perceived difficulty after each reading exercise, typically two to three times per practice session. In a comparatively free approach, Raby (2007) asked learners to record

their thoughts and experiences in journals while they completed six practice sessions.

From the perspective of more recent motivational theories, a shortcoming of all but one of the studies above is that, although they collect motivational data at different points in time, they do not present an account of how situated motivation changed with time. In this regard, the motivational trajectories presented in the study by Jauregi and colleagues stand out for clearly showing how motivation changes during the course of CALL practice.

2.4.4 Objective practice logs in the evaluations

To obtain indications of learner motivation and contextualise subjective data, CALL systems can record objective behaviour data via computer logging. By hooking into a computer program's architecture, a CALL system can log the actions a learner performs in the computer practice environment, as well as actions by the system or other learners. CALL system logs are temporally accurate and can log at a detailed level and with consistency and objectiveness.

In our selection, six of the 17 studies which evaluate situated motivation make use of system logs, with five of these studies using logs to support an analysis of learner-learner interactions in CMC practice (see Table 2.3). The study by Ushida (2005) used logs to capture learner participation in chat and forum exercises and allowed teachers to review the logs to assess the quality of students' participation. In the study by Lehtonen and Tuomainen (2003), email logs revealed how difficulties in peer interactions can demotivate learners in collaborative writing activities. Logs allowed Yang (2011) to observe how teaching assistants provided scaffolded interactions with learners in the online discussion component of their practice environment. Work by Thang and Bidmeshki (2010) used logs but limited their analysis to recording student logins and did not include details on the types of activities completed or the interactions between participants. More information on these could have provided additional evidence for students' motivation (by showing that students actively practiced after they logged in) and the lack of it also prevents links from being drawn between practice events and situated motivation. Vinther (2011) examined email logs for interactions indicative of motivation in an online language exchange. Discussion between the participants on the topic of error correction revealed that group members agreed on the usefulness of correcting errors, and that some participants received more corrections than others. This may have affected learners' experiences (e.g. Lehtonen & Tuomainen, 2003 find this in their study), but due to this study's lack of subjective motivational data, it is difficult to know how practice events impacted motivation.

In autonomous CALL practice, we found only the study by Heilman et al. (2010) which used their computer logs to check learner behaviour during vocabulary practice. Logs enabled the researchers to examine a number of behaviours, such as time spent reading a passage, or number of dictionary look-ups and to corroborate their unexpected findings regarding learners'

Table 2.3: Recording practice events in CALL evaluations investigating situated motivation.

Observation type	Description	<i>N</i>	Studies
Computer logging	Computer system automatically logs learner-system and learner-learner interactions.	6	Ushida (2005); Thang and Bidmeshki (2010); Vinther (2010); Lehtonen and Tuomainen (2003); Yang (2011); Heilman et al. (2010);
Human observations	Researchers observe practice or review video recordings of language practice and document practice events.	2	Raby (2007); Fidaoui et al. (2010)
No observations	CALL practice is treated as a ‘black box’ with no analysis of what happened during practice.	9	Chang (2007); Ushioda (2000); Chen and Brown (2011); Kissau (2012); Colpaert (2010); Tennant and Gardner (2004); Van Aacken (1999); Choi and Johnson (2005); Jauregi et al. (2012)

self-reported interest in practice. The logs also served as a way of verifying that their automated reading material recommendation system worked as intended. That so few evaluations of autonomous CALL practice employ logs is an interesting finding and suggests that this is a point which could be improved on. We return to this point later when we discuss future directions for research evaluating motivational impact.

2.4.5 Global motivation in evaluations of CALL systems

In evaluations of CALL systems, data on global motivation may provide context explaining motivational impact, practice efforts or learning gains. After providing an overview of the types of global motivational information gathered by 12 studies from our selection taking a global or combined perspective on motivation, we look briefly at 10 using this information to measure the impact of training.

Types of global motivational information

What kind of information could global motivation add to contextualise situated data? The 12 studies that include global motivation in our selection employ a variety of constructs. Socio-educational constructs such as integrative and instrumental orientation appear to be the most common, with four studies adopting Gardner's AMTB as the sole measure for examining global motivation (Choi & Johnson, 2005; Tennant & Gardner, 2004; Ushida, 2005; Wehner et al., 2011). Another type of information found almost as often is related to social cognitive learning theory. Evaluations by Kissau (2012) and M. Chang (2005); M.-M. Chang (2010) gather information on learners' self-efficacy, their beliefs regarding their ability to achieve goals they find desirable, and their tendency to self-regulate, or to reflect on their methods for learning a language and use of strategies. Two other types of global motivational information are less common in our selection. Information derived from SDT was found only in the study by Morton and Jack (2010), which placed learner's motives for L2 learning along an intrinsic-extrinsic continuum. Finally, Heilman et al. (2010) collected information on learners' personal interests, considered to be "a precursor to intrinsic motivation" and to be "topic-specific and have long-lasting value" (p. 74).

Another approach is to combine constructs from different theoretical frameworks. The combinations we observed consisted of the orientations from Gardner's AMTB and other constructs. A study by Jauregi et al. (2012) combined elements from the AMTB with linguistic self-confidence and learning goals. Van Aacken (1999) situated integrative and instrumental motives within an intrinsic-extrinsic model of motivation, showing that researchers have combined different types of global motivation information for some time.

A final observation is that the L2 Motivational Self System has yet to have a significant impact in CALL. In this review, only the work by Gleason and Suvorov (2012) adopts the framework as main theoretical basis. The study examines how participants' ability to envision themselves as proficient L2 speakers is influenced by asynchronous oral CMC practice. The previously mentioned study by J. C. Chen and Brown (2011), which we have categorised as having a situated motivational perspective, also references the framework's Ideal and Ought-to self-guides in a brief qualitative account of learners' motivational experiences. Thus, though far from dominating CALL evaluations of motivation, our analysis suggests that integrative and instrumental orientations from Gardner's AMTB remain influential constructs.

Global motivation as a measure of impact

In our selections, we find one study which elicits global motivational characteristics to contextualise situated motivation and practice behaviour (Van Aacken, 1999) while another elicits global data to personalise practice (Heilman et al., 2010). The remaining 10 studies include global motivation as a measure of practice impact in their evaluations, demonstrating that CALL views global

motivation as malleable and shaped by situational factors. For example, Wehner et al. (2011) compared motivational orientations of a group who practised in an online virtual world with a control group that did not by having them complete the AMTB once after practice. A methodological shortcoming of this approach is that we cannot be certain of the equivalence of the two groups before practice began.

In general, studies assess global motivation once before practice begins and again after it has completed. The interval between assessments is typically four weeks or more. In work by Tennant and Gardner (2004), participants trained for 10 weeks and completed a reduced version of the AMTB twice, at the fifth and tenth week. In the study by Gleason and Suvorov (2012), participants in a blended learning environment began completing asynchronous oral CMC tasks in the seventh week of their course and completed a survey on their Ideal and Ought-to L2 selves later in weeks 11 and 15. Similar semester-long practice periods can be found in the studies by Ushida (2005), who evaluates three online language courses, and by Kissau (2012), who compares the self-efficacy of a group of teacher candidates training online with a group learning in a traditional classroom, and in work by Chang (2005, 2010), which measures the impact of web-based activities designed to promote reflection on study habits on self-efficacy and other sociocognitive constructs. The two shortest periods belong to a study by Morton and Jack (2010) which measured motivational orientations for two groups studying different target languages in different sociocultural contexts after a single training session, and the pilot study by Choi and Johnson (2005), which measured AMTB-scale information before and after a five-day training period. The study by Morton and Jack (2010) is noteworthy because it demonstrates how analysis of global motivation can benefit CALL evaluations. In the study, two groups of students learning two different L2s in different locations reported different experiences in practice with the same CALL system (teaching different L2s). The findings suggest that one group had a less-positive experience (participants reported increasing levels of stress) and that this could be related to the group's comparatively lower levels of intrinsic motivation and higher levels of extrinsic motivation for learning the L2, although the study cautions that more research is required. The study by Jauregi et al. (2012), already discussed above, used a questionnaire that contained items that could also be seen as measures of global motivation, such as attitudes towards the L2 culture and motivation to learn the L2. The study demonstrates the most dynamic view of global motivation in our selection, sampling weekly over a nine-week practice period.

2.5 Future Directions

The shift in interest in recent years from global to situated perspectives has yielded theoretical frameworks which have made motivation, as an evaluation criterion, more relevant to CALL by attaching importance to the effect of

the practice environment on learners' motivation. However, our review of the literature has identified a number of gaps between the approaches that modern theories recommend and how CALL research is evaluating motivational impact at present. To address these gaps, we propose four areas for improving the assessment of situated and global motivation in future evaluations:

1. The development of computerised methods for gathering information on learners' situated motivational states and global motivational characteristics.
2. The use of computers to log learner-computer or learner-learner interaction data, for the purpose of contextualising motivational outcomes with objective records of practice behaviour.
3. Evaluations of CALL systems from the perspective of Dörnyei's L2 Motivational Self System framework, to research the role of learners' global motivational characteristics in determining practice effectiveness.
4. The development of adaptive systems that personalise instruction via student modelling technologies.

We expand on each of these points below.

2.5.1 Computerised methods for gathering situated and global motivational data

Our first suggestion for future work has to do with the observation that there appears to be very little in the way of specialised computerised instruments for L2 motivation. In this review we encountered only (1) a computerised 13-item version of Gardner's original AMTB instrument, (2) a 'thermometer' user interface for measuring motivation and anxiety (both in Tennant and Gardner, 2004) and (3) a simple Likert-style web-based questionnaire gathering learners' readings interests (Heilman et al., 2010).

Though perspectives on L2 motivation have developed over the years, instruments in CALL have remained limited to the usual tools of questionnaires and interviews. On the one hand, these instruments have been used in decades of motivation research, are well understood and have extensive experimental validation. On the other hand, just as CALL has contributed new modes of practice to language learning, the CALL community, due to its familiarity with language acquisition, pedagogy, individual differences and computer technology, would seem to be well positioned to contribute innovative computer-based instruments specialising in gathering global and situated L2 motivational data.

To illustrate why future research in this direction could be interesting, we sketch three ideas that we believe are likely to benefit evaluations of motivation in CALL. Most researchers will have already replaced pen-and-paper instruments with their computerised equivalents, bringing advantages such as online access and scalability to large groups. We think the next step is to integrate

these instruments into CALL systems, to be able to fine-tune their user interface to each particular system, but also for the ability to log learner behaviour while they interact with the instrument and to better support student modelling technology (see the sections on logging and student modelling below).

A second idea is to make more use of the adaptive capabilities of computers. Briefly, computer adaptive testing (CAT) instruments can select or prune lines of questioning according to the aspects of motivation that emerge during the completion of a questionnaire. This technology could be used to survey participants' attitudes with fewer or more relevant questions.

A third possibility, and the most ambitious one here, would be efforts to develop computerised instruments that measure motivation objectively. For example, eye-tracking technology has recently been used in CALL research to detect whether learners notice corrective feedback (Smith, 2013) and with non-CALL e-learning systems for tracking engagement and interest (Wang, Chignell, & Ishizuka, 2006), though we did not find the technology employed to research motivation in this review. Perhaps this technology could also be used in CALL evaluations to measure learner engagement in practice. In the case of CALL systems which provide L2 output practice, another interesting idea would be the application of automatic affective speech processing or sentiment analysis to speech or written text produced by the learner. Another possibility may be to measure a "person's behaviour in a standardized test situation rather than asking him/her explicitly" (Winkel & Petermann, 2007, p. 32). Examples of standardised tests could include a curriculum arrangement activity in which learners browse and assemble units to create their own language learning curriculum, or simulations and role-plays (e.g. where a learner chooses different topics in a simulated conversation, or pursues different paths in a branching story line).

2.5.2 System logs: a resource for contextualising motivational data

Based on our review, we have found that many evaluations of situated motivation treat practice as a 'black box' and do not include an analysis of practice events (Chang, 2007; Choi and Johnson, 2005; Jauregi et al., 2012; Tennant and Gardner, 2004; Van Aacken, 1999). Other studies include practice events, but in our view fall short of demonstrating the kind of contextualisation possible in a computer environment: some of the studies look at practice events through retrospective methods which are limited to what the participant can report (Chen and Brown, 2011; Colpaert, 2010; Kissau, 2012; Ushioda, 2000); others use a system's logging facility but do not report in sufficient detail on the activities of learners (Thang and Bidmeshki, 2010), or rely on the logs to measure motivation and do not complement this data with more direct methods such as interviews or questionnaires (Vinther, 2011). This leaves a small number which demonstrate the use of logging to contextualise situated motivation, with one group using human observations to record practice behaviours (Fidaoui et

al., 2010; Raby, 2007), and a final group using computer logs (Heilman et al., 2010; Lehtonen and Tuomainen, 2003; Ushida, 2005; Yang, 2011).

Given that logging modules are frequently available in CALL environments, we argue that another important way of reducing the gap between theory and CALL evaluations is to do more with the logs. Here we present three ideas for how CALL evaluations of motivation can better exploit log data (summarised in Table 2.4).

Table 2.4: Three ideas for how evaluations of motivation can better exploit system log data.

Employ system logs to	Procedure
Evaluate validity of learner self-report data	(1) Log learner behaviour while working with self-report instruments. (2) Search log files for behaviour that suggests comprehension and (dis) engagement, such as reviewing instructions, accessing help items, skipping optional exercises, or unusually short or long completion times.
Cross-check what learners report with what they actually do	(1) Log how long learners practice, when they practice, what they practice, and how well they perform (2) Compare their practice behaviour with their motivational experiences reported in questionnaires and interviews.
Search for links between practice events and motivational outcomes	(1) When learners refer to events or patterns of events that impact motivation, use logs to access the details of the events. (2) Observe the presence or frequency of different practice events or patterns of events and investigate the correlation between these and motivational experiences

A first idea is that logs can benefit evaluations by providing a way of increasing the validity of self-report data. The suggestion is that we record learners' behaviour while they self-report; later, we search in the logs for behaviour that casts doubt on the validity of the corresponding self-report data and then treat the data appropriately (e.g. by excluding it from analysis). A few examples of behaviour we might search for are participants who spend very little time reading questionnaire instructions, skip optional practice activities, or take an unrealistically long or short time to complete individual items or the questionnaire. The same process could be used to lend support to data that might otherwise be considered doubtful. In this case, we would look for

behaviours that suggest comprehension and engagement, such as accessing links to on-screen help items or tracing questionnaire items with the mouse. Additionally, behavioural data such as these could be combined with searches for patterns in the values learners' report, such as alternating use of the extreme poles of a Likert scale, to identify data with questionable validity.

Our second suggestion is most relevant for evaluations of relatively free practice where learners have control over what they practice, how long they practice, and when they practice (e.g. Raby, 2007; Ushida, 2005). The idea is to use the log data from language practice to cross-check what learners report about their motivation with what they actually do. A known weakness of self-report instruments is that they cannot detect when a learners' reports have been influenced by external factors such as social desirability or concerns about confidentiality. This issue can be partially addressed in CALL evaluations by examining, for example, the actual study behaviours of self-described highly motivated persons. We can check if time on task matches with learners' reports of motivation. Similarly, if learners report that they feel more confident in their abilities to do well in practice, then one can look in the logs to see if their performance has improved, or at least not worsened.

Our third and final suggestion is relevant for the relatively uncontrolled type of practice described above, but also for evaluations of practice in an experimental setting where learners do not control their practice activities or time on task (e.g. Heilman et al., 2010; Morton and Jack, 2010). The idea is to attempt to use the logs to identify practice events or patterns of events that influence motivational outcomes. Here, we sketch a procedure for how this might be done. The method we suggest has two steps:

1. In a first step, we would examine the interactions which take place during the practice period and attempt to identify those that influence motivation (or other related dependent variables of interest).
2. In a second step, we would characterise the practice periods using quantitative descriptors of interaction tendencies and use statistical analysis techniques to search for links between interactional tendencies and motivational outcomes.

How we go about the first step of identifying interactions that influence motivation could take at least two approaches. One approach would be to consider which interactions are possible, given the practice task and program user interface, and make informed predictions on which interaction patterns might affect situated motivation (e.g. del Soldato & du Boulay, 1995). Examples of patterns we might look for include interactions which suggest learning, or on the other hand, a task that is too easy or difficult. We might also look for patterns suggesting disinterest in particular course materials or practice targeting particular linguistic skills, such as extremely slow or fast response times (Beck, 2005) or non-cooperative behaviours. Long page views and back and forth behaviour between pages coupled with poor task performance could indicate problems with task comprehension.

A second option likely to provide useful insights into the link between practice events and motivation would be an in-depth qualitative approach. Using this method, the researcher could use the interview techniques (e.g. Chen and Brown, 2011) or focus groups (e.g. Colpaert, 2010) that CALL already uses to ask learners or teachers about moments during practice where they felt motivated, confident or frustrated. In a second step, interviewers could ask learners to reflect on the events that led them to feel this way. Finally, these events could then be located in the logs to obtain the details of the interaction. An issue with this approach is that “our understanding of how motivation evolves...is limited to what our research participant can usefully tell us” (Ushioda, 2012, p. 68), but an innovative method taken up by some scholars may offer help. Briefly, the method is to use video recordings to help learners remember their practice experience. Investigating the feasibility of diagnosing motivational state in their vocabulary tutor, de Vicente and Pain (2002) investigated an approach by asking teachers to rate student affect and motivation based only on recordings of on-screen practice behaviour. In another study, MacIntyre and Legatto (2011) asked students to review a video of themselves speaking in a context similar to an oral examination and report their levels of willingness to communicate. Though these studies were concerned mainly with accurately diagnosing motivational state during practice, in our view these techniques could be adapted to ask students to explain what factors influenced their motivational state at the time.

Once we have identified the motivationally relevant interactions, a second step would be to employ them as quantitative descriptors of practice sessions and to search for links between these and motivation. Due to limited space we can provide only a short description, but a relatively straightforward approach would be to profile the practice sessions using the frequency of different relevant interactions obtained in the previous step. A correlational analysis could then be employed to explore the relationships between interaction patterns and motivational trajectories. Such a correlational analysis could be performed on a group dataset containing multiple participants, as in quantitative approaches, or on an individual’s dataset, in keeping with qualitative approaches. The aim of the former would be to search for common tendencies of certain types of interactions to yield certain types of motivational outcomes, while the latter approach could be helpful for explaining the trajectory of individuals.

2.5.3 Investigations into the role of global motivational data from the perspective of the L2 Motivational Self System

Another way we could reduce the gap is by attempting evaluations of motivation from the perspective of the L2 Motivational Self System. Most global theories of L2 motivation are represented in the CALL literature, but we saw that this theoretical framework has yet to make a significant impact. However, there are compelling reasons for CALL studies assessing global motivation to do so from

this perspective. First, the framework has strong theoretical foundations in motivational psychology, with demonstrations of validity and practical utility in a variety of domains such as career planning, education and many others (Dunkel & Kerpelman, 2006). Second, though relatively young in SLA, the perspective has already received substantial experimental validation in large quantitative studies by Dörnyei and other researchers (Dörnyei, 2009a). Finally, in viewing motivation as a learner-internal process involving the comparison of personally desired (or undesired) futures, the framework generalises to different sociocultural contexts and is well-suited as a basis for personalised instruction in CALL systems (see section below).

New research on global motivation which adopted this theoretical basis could search for relationships between a learner's L2 self-guides and situated motivation, practice behaviour and changes in L2 proficiency. This type of research would be an important first step in a larger research effort aimed at assessing the impact of personalised CALL practice, by first establishing that a learner's global motivational characteristics (i.e. their self-guides) are related to their situated motivation or practice behaviour and by capturing the associated motivational or behavioural patterns.

2.5.4 Personalised practice through student modelling technology

Personalisation (or individualisation', or 'adaptation') of instruction has received much attention in a sub-area of CALL known as intelligent CALL (ICALL) and intelligent tutoring system (ITS) research. Research of this type makes use of a technology called student modelling to store information about individual learner differences in computer programs. Generally, student models are associated with a user account and are loaded after a student logs in to a system. Such student models can inform other computer program modules, such as those responsible for the selection of practice materials. A major topic of research in ICALL has been the use of student modelling technology to capture individual differences in proficiency and adapt aspects of instruction such as feedback behaviour (for a review, see Heift and Schulze, 2007). However, adaptation based on motivation-related individual differences is, at present, very rare, with the work by Heilman et al. (2010) being the only system in our review that makes use of a student model for adapting to motivation-related individual differences, and perhaps the first CALL study to assess the impact of an automatic motivational strategy. Along the lines of Ushioda's (2009a; 2011; 2012) views on the personalisation of instruction, one way of bridging the gap is to investigate how global and situated motivational data can be used as a resource for adapting instruction to individual learners.

As we explore the possibility of automatically gathering information on learners' goals and interests related to L2 learning and adapting computerised instruction, we also have to be realistic about the capabilities of current technologies. A person's goals are various, rich in experience, detailed and

often unique, and it would be impossible for ICALL or ITS technology to capture them in their detail with existing technology. Because of limits in technologies that process written language and speech, the expression of individuals' goals and interests in natural language, such as in open-ended questionnaires or spoken interviews, are likely to be beyond the state of the art at present. Thus, while research into the possibilities of current technologies for gathering motivational information is potentially a new and interesting area, it is definitely not the case that progress in CALL motivation research is dependent on these developments. On the contrary, this review has shown that qualitative analysis of open questionnaire (Lehtonen and Tuomainen, 2003; Ushioda, 2000), interview (e.g. Chen, 2011; Kissau, 2012; Van Aacken, 1999) and group discussion data (e.g. Colpaert, 2010) can provide important insights into learners' global and situated motivation. It follows that the field should stay up to date with developments in non-automatic methods while investigating the possibilities with technology in parallel.

2.6 Conclusion

In recent years, the number of computer programs that serve learners as L2 practice environments has grown noticeably. At present, there are many options, from authentic target language environments intended originally for native speakers, to environments designed for language learners, the latter providing autonomous and CMC practice activities, in synchronous and asynchronous modes, and offering form-focused and meaning-focused activities. As these practice environments become a standard part of language learning, it becomes increasingly important that we have methods for selecting the right computer applications, and that these methods include an assessment of motivational impact.

In completing this review we have found that there is a gap between how recent theories recommend evaluating motivation and how CALL studies which evaluate motivational impact operate. To close the gap, new CALL studies will need to employ theories of situated motivation, monitor changes in motivation over time, and contextualise situated motivation data by (1) exploiting computer logging to record practice events and (2) gathering global motivational data. Evaluations with these characteristics have the potential to provide a more detailed account of motivation in practice. They can present a dynamic picture of learners' general appraisal of a system in line with current dynamic systems perspectives of SLA (see, for example, de Bot, Lowie, & Verspoor, 2007) that highlight the temporal dimension and the importance of context. L2 motivation theory also appears to be heading in this direction (Dörnyei & Ushioda, 2011).

In this paper, we have attempted to provide a number of suggestions for future research that evaluates the motivational impact of CALL systems. Admittedly, the methods we have suggested have been about harnessing

technology to study L2 motivation, and as a result, are mainly quantitative. However, as we saw in this review, much CALL practice involves social interactions between learners and teachers. Qualitative methods seem particularly well suited for understanding how social interactions can impact motivation. Thus, an assessment of the state of qualitative methods in CALL studies would complement our review and is in our view an interesting and important area for future work. Our aim with the suggestions we have provided is to stimulate the use of evaluation methods that provide a more fine-grained account of practice that perhaps do a better job of explaining the process leading to motivational and L2 learning outcomes. An enhanced understanding of motivational processes has the potential to be particularly beneficial in CALL evaluations, because it goes one step further than determining the pedagogical effectiveness of a CALL system to also document the chain of events that lead to an outcome, and thereby contribute avenues for improving CALL instruction.

CHAPTER 3

Learner affect in computerised L2 oral grammar practice with corrective feedback

This chapter has been accepted as

Bodnar, S., Cucchiarini, C., Penning de Vries, B.W.F., Strik, H., & van Hout, R.W.N.M. Learner affect in computerised L2 oral grammar practice with corrective feedback (to *Computer Assisted Language Learning*)

Acknowledgements

We are indebted to three anonymous reviewers for their comments on an earlier version of this chapter. This work is part of the research program Feedback and the Acquisition of Syntax in Oral Proficiency (FASOP) NWO-Dossier Number 360-75-010, which is funded by the Netherlands Organisation for Scientific Research (NWO).

Abstract

Although corrective feedback (CF) has received much interest in the second language acquisition literature, relatively little research has investigated the relationship between CF and learner affect in concrete practice situations. The present study adopts a situated perspective to investigate learners' affective states and practice behaviour in a novel context: oral grammar practice with a computer-assisted language learning (CALL) system employing automatic speech recognition technology (ASR) to analyse learners' speech and provide feedback. Thirty one adult learners of Dutch practiced with this system in one of two conditions: the no feedback condition (NOCF), and the feedback condition (CF) which provided immediate CF through ASR. Despite concerns that CF can elicit negative affective reactions and although practice with feedback forced learners to reformulate more often, CF did not appear to have a negative impact. Our analysis finds no significant differences between the NOCF and CF groups. Within the CF group, learners who received more CF tended to report less self-efficacy. These findings suggest that ASR-enabled CALL systems may be suitable environments for oral grammar practice where CF on oral productions can be provided without negative affective reactions.

Keywords: corrective feedback, affect, grammar, computer assisted language learning

3.1 Introduction

The ability to speak a second language (L2) proficiently is a skill that is typically highly desired by learners and valuable information about how acquiring this skill can be facilitated can be obtained through research that evaluates the instructional practices adopted. One instructional practice that is thought to be helpful, but where additional research is needed is corrective feedback (CF). Learners inevitably make errors on their way to learning a language and the question is how these should be dealt with when they occur (Lyster & Ranta, 1997). While CF has long been a controversial issue, evidence from recent meta-studies now points towards a facilitative effect on L2 proficiency development (Lyster, Saito, & Sato, 2013), but many details concerning how CF impacts L2 learning and motivation remain to be worked out.

Discussions of the potential for CF to have a negative impact on learner affect and motivation can be found in many places in the SLA literature (Agudo, 2013; Ellis, 2009; Kartchava, 2012; Krashen, 1982; Loewen et al., 2009; Lyster et al., 2013; Truscott, 1999; Vásquez & Harvey, 2010), and to some extent in the CALL literature (Schulze, 2003), but few studies have investigated this relationship (Ellis, 2010).

Emotion-related dimensions have been linked to important functions in learning (Fredrickson, 2001; Pekrun & Linnenbrink-Garcia, 2012) and are seen as a source of motivation (P. D. MacIntyre, 2002). In considering whether feedback is effective, it is important to remember that learners are emotional beings as well as cognitive ones and that these two components interact with any learning condition to influence its outcome (Gardner, 1985b; Pekrun & Linnenbrink-Garcia, 2012). This is especially relevant to oral L2 practice, as research has shown that learners tend to be anxious in oral L2 communication in front of instructors or their peers (e.g. Horwitz, Horwitz, & Cope, 1986). CF provided in the context of oral L2 practice in the classroom, while likely beneficial from a cognitive point of view, could be unpleasant for some learners and may render them less willing to speak (Agudo, 2013). Research into the pedagogical effects of CF should therefore also address affective dimensions of the learner, such as emotion and motivation. From this point of view, computer environments that provide L2 speaking practice are particularly interesting because they are often less threatening (Warschauer, 1996).

The present study complements previous research on the effect of computerised oral L2 grammar practice on L2 proficiency (Penning de Vries et al., 2014b; Penning de Vries, Cucchiaroni, Bodnar, Strik, & van Hout, *in press*) with an evaluation of the practice environment from an affective point of view. In line with recent perspectives on L2 motivation (e.g. Ushioda, 2012; see also Ushioda & Dörnyei, 2012), we adopt a situated approach to studying affective experiences (cf Turner & Trucano, 2014) in computerised spoken grammar practice which complements previous CALL and affect research (e.g. J. C. Chen & Brown, 2011; Heilman, Collins-Thompson, Eskenazi, Juffs, &

Wilson, 2010; Jauregi, de Graaff, van den Bergh, & Kriz, 2012) in two ways: First, the method adopted in the present study provides a dynamic description of learners' affective states while they practice with the system. Second, we employ a logging module to record student and system behaviour, making it possible to look for relationships between affective states and different types of events that occur during practice (cf Ushioda, 2012).

In what follows, we begin with a review of research into the interaction of CF and learner affect. We then present the approach, the research questions and the methodology in our study. Finally, results are presented and discussed in relation to the literature and pedagogical implications are suggested.

3.2 Background

In the SLA literature, affective variables can be broadly defined as “those emotionally relevant characteristics of the individual that influence how she/he will respond to any situation” (Gardner, 1985b, p. 1). From the point of view of evaluating instructional practices in computerised learning environments, we have found it useful to differentiate between affective variables according to their sensitivity to context: On the one hand are affective characteristics of a person that are considered to be stable and context-independent. For example, attitudes are affective experiences that are relatively long in duration and have a less direct relationship with behaviour (Shuman & Scherer, 2014). A second example is the trait concept, which is “not an affective experience itself, but a tendency to experience a mood or emotion” (ibid, p. 18). In the L2 literature, trait anxiety (Horwitz, 2010), integrative or instrumental motivational orientations (Gardner, 1985b), or global judgments of foreign language learning enjoyment (Dewaele & MacIntyre, 2014) can be seen as affective characteristics that are relatively context-independent.

On the other hand are affective characteristics of a person that are more transient and tied to the immediate learning environment. Emotions, including such states as happiness, embarrassment and boredom (Garrett & Young, 2009), fall into this category, as do moods, which can be differentiated from emotions by the fact that they are not directed towards a specific target and that they are longer lasting (Shuman & Scherer, 2014). In emotions research, emotions and moods are collectively referred to as affect (Pekrun & Linnenbrink-Garcia, 2012). As the present study is concerned with the more transient affective phenomena potentially impacted by the learning context, we use the terms *affect* and *affective state* in this sense, to refer to subjective feelings that are experienced in a particular language learning environment, including emotional states and other transient complex states such as “feelings of self-confidence, feeling willing to communicate, or feeling anxious” (MacIntyre & Gregersen, 2012, p. 103) or state motivation (Tennant & Gardner, 2004; Tremblay, Goldberg, & Gardner, 1995).

The relation between affect and CF is an area that clearly concerns teachers

(Vásquez & Harvey, 2010) as well as SLA and CALL researchers (Ellis, 2009; Krashen, 1982; Schulze, 2003; Truscott, 1999), though little research has investigated the connection between affect, as defined above, and feedback, especially in an oral practice context (Ellis, 2010). Other directions CF research has taken include investigating how more stable affective characteristics such as trait anxiety influence the effectiveness of CF in classroom contexts (Sheen, 2008; see also Dasse-Askildson, 2008; Uzum, 2011) and surveys of learners' general beliefs, attitudes or perceptions about feedback (Agudo, 2013; Cathcart & Olsen, 1976; Kartchava & Ammar, 2013; Loewen et al., 2009; Plonsky & Mills, 2006).

The latter group of studies have produced findings which together suggest that learners' perceptions of CF are mixed. Kartchava and Ammar (2013) found that learners valued CF, but also agreed with statements that CF causes anxiety or that oral corrections can lead to negative attitude towards studying an L2. Agudo (2013) found that most students believe CF is effective, but that a large number also have negative affective responses. Loewen et al. (2009), eliciting learners' beliefs and perceptions of grammar practice and CF, found that participants viewed CF somewhat negatively.

Two classroom studies are worth a closer look here because of the discrepancy they seem to show between learners' generally stated beliefs and their beliefs regarding CF in a concrete learning context. Cathcart and Olsen (1976) surveyed students' perceptions of CF in oral contexts and found that students had positive attitudes towards CF, not only in drills, but during all types of practice. However, in a follow up exercise in which a teacher corrected all errors in their oral productions, students acknowledged the difficulty of having a coherent conversation while being constantly corrected. Plonsky and Mills (2006) investigated the gap between students' and teachers' beliefs regarding the potential for CF to discourage learners. Prior to treatment, learners seemed to recognise the potential for discouragement less than the teacher, but after viewing and discussing different types of CF practices with teachers they acknowledged that CF could have a negative impact and that not all errors require CF. These findings provide some support for Ushioda's (2012) recommendation to attempt more in situ analyses of motivation, which seems equally applicable to research on other affective characteristics.

In CALL research, the relationship between CF and affect has received limited attention (but see Schulze, 2003 for a discussion of its importance). The existing studies have the characteristic of researching affect in a concrete learning context. In the context of an immersive self-study dialogue game for English pragmatics, Cornillie, Clarebout, and Desmet (2012) investigated the links between the perceived usefulness of CF and learners' intrinsic motivation for learning English, game experience and perceived in-game competence. The study examined three different types of CF and found that an explicit, non-immersive CF was most preferable, despite the realistic communication-oriented type of practice, and that learners who perceived the explicit CF as more useful tended to report more positive game experiences and stronger

feelings of competence.

Drawing on skill acquisition theory, Cornillie and Desmet (2013) investigated if CF enhanced with game design features could increase motivation in English grammar drills. Game players routinely experience failure, yet by adding elements of fantasy and vividness, so-called positive failure feedback (PFF) can elicit positive emotional responses. A preliminary analysis suggests that PFF may have contributed to learners' feelings of immersion while practising, although interviews with participants revealed that some found PFF distracting or frustrating. A correlational analysis suggests that learners who felt more immersed and competent tended to report more enjoyment and were more willing to practice. No evidence was found for a link between PFF and perceived in-game competence.

Summarising, CF has been an important topic in SLA and CALL research for some time, but the question raised by Krashen (1982) and Truscott (1999) of whether or not CF has a negative impact on learner affect has been investigated only at a general level, with some evidence showing that learners' generally stated attitudes may not always apply to concrete learning contexts. Thus, research targeting learners' affective state in a concrete learning context would complement the existing research, especially in the case of oral L2 practice where anxiety tends to be high.

3.3 The present study

The present study adopts a situated perspective to investigate the impact of CF on learner affect in oral grammar practice with a CALL system. Three characteristics of our approach contribute to its situated character: First, the study focuses on affect in a concrete learning environment. Second, it targets affective states that learners' experience during practice with a methodology that asks learners to report their affect during practice via a short questionnaire. Third, we attempt to identify practice behaviours that are related to affect by searching for relationships between affect and system-learner behaviour logs. An important aspect of our strategy in measuring in-practice affective states was to minimise interruption by limiting the questionnaire to a small number of constructs, a difficult task given the multi-faceted nature of a learner's affective practice experience. While theoretical frameworks such as Keller's ARCS (1987) and Csikszentmihalyi's flow theory (Csikszentmihalyi, 1975; see also Egbert, 2003) provide constructs to investigate, our approach was to survey the SLA literature and select constructs which could plausibly be expected to be sensitive to the learning environment and CF.

As a first construct, we chose enjoyment with the aim of describing learners' emotional response to practice with the system. Enjoyment is a positive emotional state associated with feelings of satisfaction and pleasure arising from participation in an activity (Ainley & Hidi, 2014) and reflected when learners express that they 'like' or 'enjoy' an activity, or find it 'fun' or 'nice'

(Dewaele and MacIntyre, 2014). Importantly, positive affective experiences such as enjoyment are thought to foster positive learning outcomes (Fredrickson, 2001; Pekrun & Linnenbrink-Garcia, 2012). In the SLA literature, we have seen that CF is sometimes associated with negative affective experiences in the classroom, and therefore there is a potential for CF to negatively impact learner enjoyment. However, computerised self-study environments may be a context where this concern does not apply and are therefore worth investigating.

For the second construct, we were interested in a motivational component which recorded the learners' willingness to practice with the system, similar to dynamic approach-avoidance motivation (P. MacIntyre & Serroul, 2015) and willingness to communicate (WTC) (MacIntyre & Legatto, 2011). Previous CALL research has demonstrated that WTC can be sensitive to the learning context (Jauregi et al., 2012). An important difference between studies using WTC, which "represents the intention to speak or to remain silent" in the context of "conversation with a specific person at a specific moment" (MacIntyre & Legatto, 2011, p. 150), and the present study is the context: the target behaviour is not communication between persons, but a willingness to engage in accuracy-focused oral L2 production in a computerised self-study context.

For the third and final construct, we were interested in the feelings of competence a learner might experience during practice. Ushioda (2007) suggests that learners have a need to "be competent and effective in their interactions with the world" (p. 78) and that teachers' responsibilities also include providing scaffolding to maintain motivation and control frustration. In this connection, we decided to measure self-efficacy, the beliefs one has about one's ability to accomplish a specific task (Bandura, 1978). Following Narciss (2004), we assumed that CF, by providing opportunities to notice and correct one's errors, can provide a sense of learning and foster self-efficacy. Tracking learner's self-efficacy can indicate whether or not learners felt more capable over time, which in turn can provide an indication of CF effectiveness.

The research questions for the study were the following:

1. Are there differences in learners' reported affective states for those who practice without CF compared to those who practice and receive immediate CF? (RQ1)
2. Are there relationships between learner affect and learner behaviour? (RQ2)

3.4 Methods

3.4.1 Participants

Participants were 31 Dutch L2 learners (males = 15; females = 16, age ranged from 18 to 62, mean age of 29) at the A1 or A2 level of the Common European

Framework (CEF) who practiced speaking Dutch using a computer system. They received 20 Euros for participating. Most participants reported having attended post-secondary school (84 percent), while four others reported four years or more of public school (13 percent) and one person reported having completed elementary school only. Learners had lived in the Netherlands between two months and four years ($M = 12$ months). They had been studying Dutch between one and 48 months ($M = 6$). In total there were 18 different L1s. The subjects also reported having a variety of L2s, with the most common being English (49%), French (14%) and German (11%). At the time of the experiment they were attending classroom lessons twice a week for a total of three hours per week.

3.4.2 Data collection

Dutch L2 Oral Grammar Practice with an ASR-based CALL system

For L2 learners of Dutch, certain aspects of Dutch word order can be difficult to master. One example is verb second (V2) (Jordens, 1988), which requires inverting the position of the subject and verb when a phrase begins with an adverbial constituent. The CALL system we developed, called GREET, provides exercises targeting this and other aspects of Dutch word order. Learners watch video clips and complete oral quizzes in which they respond to questions about the clips by recording their answers (see Figure 3.1). An innovative aspect of the system is that it uses automatic speech recognition (ASR) to analyse learner speech and provide immediate CF of a prompt type (for a detailed description see Penning de Vries et al., 2014b). First, the system displays a feedback message formulated as ‘That was incorrect. Please try again’. On second and later attempts, the system also guides the learner towards the correct word ordering by incrementally revealing the next word block in the sequence. Learners are required to produce the correct utterance before moving on to the next question¹.

In-practice questionnaire

A key component we added for this experiment is a questionnaire designed to tap into learners’ affective states during practice (see Figure 3.2). Similar to Moos and Azevedo (2008), we intended to have learners complete this questionnaire at regular intervals. As a result, the questionnaire was designed to be as minimally interruptive to practice as possible (cf. Dela Rosa & Eskenazi, 2011). Our strategy in meeting this condition was to limit the number of items and to reuse the same layout for the questionnaire so that learners could complete it quickly.

¹A skip button was also added to allow learners to move to the next question, to handle the rare event that their utterances were repeatedly misclassified as incorrect. The button is disabled until the learner completes three attempts.

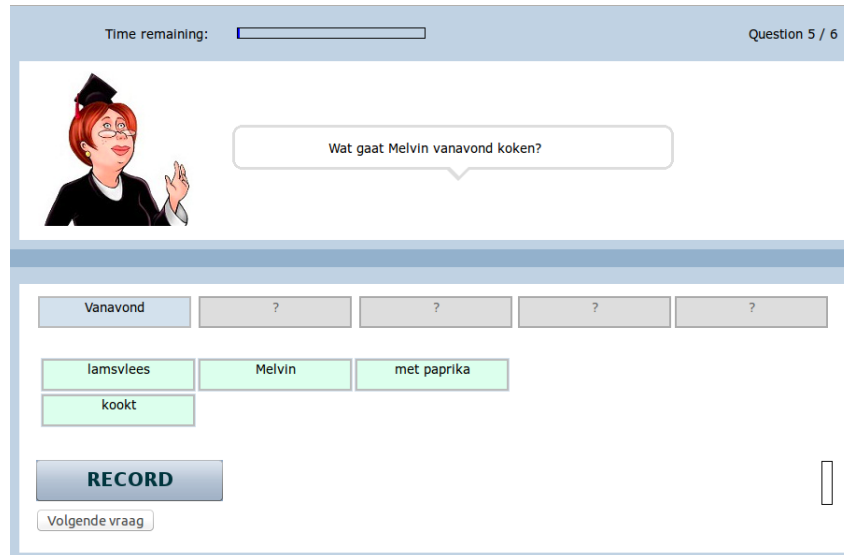


Figure 3.1: A screenshot of the ‘GREET’ CALL application. In this example, the tutor asks “What will Melvin cook tonight?”. To answer, learners compose an utterance using the fixed prompt (‘Vanavond’) and word blocks in the bottom half of the screen. A correct order is given by the block sequence “Vanavond | kookt | Melvin | lamsvlees | met paprika” (Tonight cooks Melvin lamb with peppers); a typical V2 inversion error is “Vanavond | Melvin | kookt | lamsvlees | met paprika” (Tonight Melvin cooks lamb with peppers).

In developing the questionnaire, attention was given to choosing formulations specific to the learning environment, in line with Dela Rosa and Eskenazi (2011). The three constructs were realised as five-point semantic differential items (see Figure 3.2):

- Enjoyment: *Oefenen met GREET is ... (Helemaal niet leuk - Heel leuk)*
Practising with GREET is ... (Not nice at all - Very nice)²
- Willingness to practice: *Op dit moment wil ik ... (Iets anders doen - Verder oefenen met GREET)*
At this moment I want to ... (Do something else - Continue practising with GREET)
- Self-efficacy: *De volgende vragen zullen ... zijn voor mij (Moelijk [sic] -*

²The first time this item was presented, the slightly different wording of *Ik denk dat het oefenen met GREET ... zal zijn (Helemaal niet leuk - Heel leuk)*. I think that the exercise with GREET will be ... to reflect that the user had not yet practiced with the system

Makkelijk)

The next questions will be ... for me (Difficult - Easy)

The screenshot displays a questionnaire interface with three distinct sections, each containing a visual aid, a rating scale, and two anchor labels.

- Section 1: Oefenen met GREET is _____ [English]**
 - Visual aid: A baby (left) and a monkey (right).
 - Labels: "Helemaal niet leuk [English]" (left) and "Heel leuk [English]" (right).
 - Rating scale: Five empty boxes with a mouse cursor icon positioned over the third box from the left.
- Section 2: Op dit moment wil ik _____ [English]**
 - Visual aid: Silhouettes of people playing (left) and a book titled "DUTCH" (right).
 - Labels: "Iets anders doen [English]" (left) and "Verder oefenen met GREET [English]" (right).
 - Rating scale: Five empty boxes with a mouse cursor icon positioned over the fourth box from the left.
- Section 3: De volgende vragen zullen _____ zijn voor mij. [English]**
 - Visual aid: A person looking confused with a question mark (left) and a person smiling with hands on hips (right).
 - Labels: "Moelijk [English]" (left) and "Makkelijk [English]" (right).
 - Rating scale: Five empty boxes with a mouse cursor icon positioned over the second box from the left.

Figure 3.2: A screenshot of the in-practice questionnaire. The in-practice questionnaire (see text for translations). Learners indicate their current enjoyment, willingness to practice, and self-efficacy by marking one of the empty boxes with their mouse.

Computer practice behaviour logs

The GREET system is equipped with a logging library that records learner and system behaviour. Examples of some of the events the system records are when a learner views a video or question page, when the learner records an answer for a question, and when the system determines whether or not a recorded answer is correct. These and other types of events can be used to define quantitative measures that describe the characteristics of a practice session. For example, we can calculate number of questions completed, number of attempts made and the number of questions answered correctly.

3.4.3 Experimental procedure

For experimental purposes, we configured two versions of the CALL system: the CF version used ASR to analyse learners' recordings and provide immediate CF as described above. The no-feedback version (NOCF) substituted CF with a neutral message stating that the recording had been saved and prompted the learner to choose between recording another attempt or moving to the next question. Before practice, the NOCF system told participants that their scores would be evaluated at a later time. All other aspects of practice are identical.

Participants were randomly assigned to use either the CF system or NOCF system. Each participant completed two 90-minute sessions on two separate days (typically separated by a week). In session 1, learners completed a background questionnaire and two proficiency tests before practicing. In session 2, they began practice immediately and then completed two proficiency tests and a general evaluation questionnaire.

The practice component in each session lasted 45 minutes. To control for time on task, the system was configured to return learners who completed the practice material early (in less than 45 minutes) to the start of the material, where they were asked to continue practising.

To measure learners' affect during practice, we divided each practice session into three 15-minute segments, which we refer to as micro sessions (for a total of 6 micro sessions). Before beginning the first micro session, participants completed the in-practice questionnaire for the first time. They then practiced with the system for 15 minutes, three times in a row in each session. At the end of each micro session, they completed the in-practice questionnaire, for a total of four data points per session and eight overall. In the present study, two subjects who did not complete all eight in-practice questionnaires are excluded from the analysis below.

3.4.4 Data analysis

Enjoyment, willingness to practice and self-efficacy

To detect whether there were significant changes over time and significant differences between the practice conditions, we conducted a repeated measures ANOVA test, with the factors *group* (NOCF or CF condition), *session* (session 1 or 2) and *time* (the within-session point in time at which the data was collected).

Practice behaviour

Similar to the affective measures analysis, a repeated measures ANOVA test with factors *group*, *session* and *time* is used to search for differences in learner behaviour between the groups and for changes over time. While the variety of events recorded in the logs meant that there were many possibilities for defining practice measures, the present study limits its scope to two behaviours, chosen

to illustrate the differences between the two practice conditions and because of their relationship with the affective measures (see below). First, the ratio of attempts made per question illustrates the effect of the CF operationalisation on question-answering behaviour. In the CF case, this measure also indicates performance³. Second, the proportion of questions answered correctly on first try (FirstTryOKs) serves as a measure of performance that applies equally to both conditions. Using these measures, we processed the logs to obtain quantitative descriptors of each micro session (three per session, six in total) for the ANOVA tests.

Relationships between practice behaviour and affect

A correlational analysis is used to investigate the relationships between practice behaviour and affect. The analysis computes correlations at a global level, on practice behaviour and affective data averaged across the whole experiment (including the initial questionnaire data taken before practice began), with each participant contributing one data point. If a correlation was found to be significant, we also inspected whether it existed at the session and micro session levels. The session-level correlations were computed on the mean of the measures for each session, with each participant contributing one data point per session. The micro session level correlations (six in total) were computed using the affective value provided by the learner after the practice session.

3.5 Results

3.5.1 Changes in enjoyment, willingness to practice and self-efficacy over time

Preliminary analyses pointed to significant changes in the affective variables over the course of practice, but did not produce significant group effects or group-session or group-time interactions. Because of the complete absence of group-related effects, we proceeded by analysing the changes over time by including only *session* and *time* as independent within-subject variables (see Figure 3.3).

Enjoyment

Analysis of enjoyment returned a significant effect for time, $F(3, 84) = 3.465$, $p = .050$, $\eta_p^2 = .110$, a near-significant effect for session, $F(1, 28) = 3.332$, $p = .079$, $\eta_p^2 = .106$, and no effect for the interaction between session and time ($F < 1$). Posthoc analysis of the time effect (using the Sidak procedure) reveals

³The CF condition required learners to produce the correct utterance before advancing to the next question. In the NOCF condition, attempts per question does not necessarily reflect performance (learners were free to attempt a question multiple times and may make multiple correct attempts).

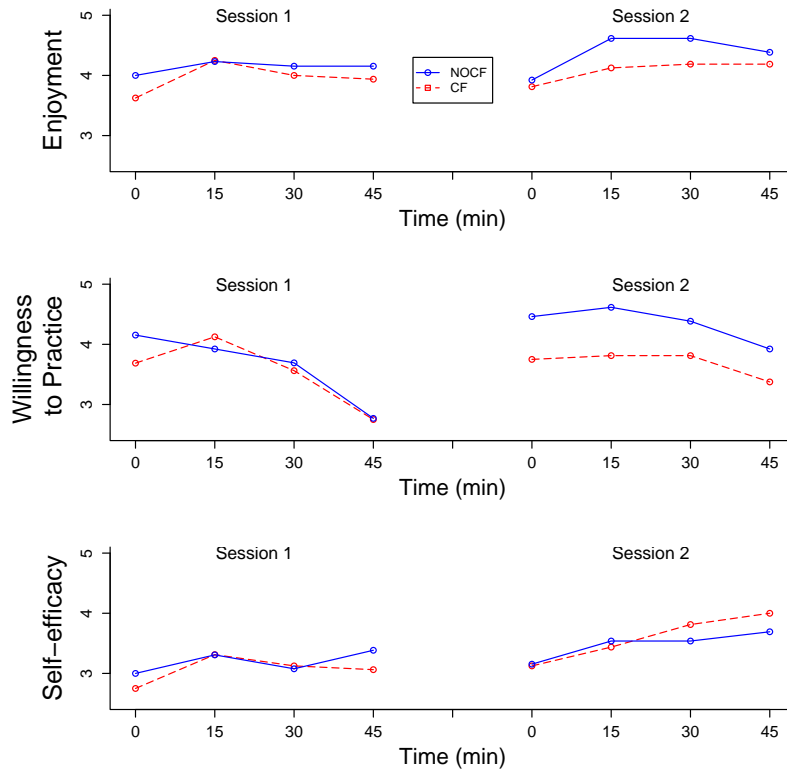


Figure 3.3: Changes in affective state for the NOCF and CF groups. Within-session affect is sampled once before practice begins and then approximately every 15 minutes over a 45-minute period.

that initial micro sessions ($M = 3.83$) score significantly lower than the second and third micro sessions (means respectively 4.29, 4.22, 4.16). Learners seem to enjoy the training sessions more than they expected at the start, both in sessions 1 and 2. Although the factor session was not significant, the mean scores in session 2 were higher ($M = 4.22$) than in session 1 ($M = 4.03$), indicating that the scores did not decrease in any way.

Willingness to practice

For willingness to practice, analysis returns significant effects for session, $F(1,28) = 7.234$, $p = .012$, $\eta_p^2 = .205$, and for time, $F(3,84) = 9.727$, $p = .000$, $\eta_p^2 = .258$, whereas their interaction is not significant, $F(3,84) = 2.663$, $p = .079$, $\eta_p^2 = .087$. The session effect is straightforward, indicating that

scores rise from the first session ($M = 3.58$) to the second ($M = 3.98$), and implies that learners remained willing to practice. The time effect indicates a falling tendency as practice time increases within the same session. Post hoc analysis (Sidak procedure) points to the fourth and final micro sessions as being significantly lower ($M = 3.19$) than the first ($M = 3.98$), second ($M = 4.10$), or third ($M = 3.84$) micro sessions, no other pair-wise differences being significant. The analysis suggests a within-session falling trend, though learners became willing to use the system again at the beginning of session 2.

Self-efficacy

Analysis of self-efficacy finds significant effects for session, $F(1,28) = 4.222$, $p = .045$, $\eta_p^2 = .136$, and for time, $F(3,84) = 5.906$, $p = .006$, $\eta_p^2 = .174$, while their interaction is not significant, $F(3,84) = 1.307$, $p = .279$, $\eta_p^2 = .045$. The session effect indicates that self-efficacy increases from session 1 ($M = 3.12$) to session 2 ($M = 3.54$). For the time effect, post-hoc analysis (Sidak) points to significant differences between the initial micro sessions ($M = 3.0$) and the final micro sessions ($M = 3.53$), suggesting a rising trend within the sessions.

3.5.2 Practice behaviour in the NOCF and CF conditions

Beginning with attempts per question, analysis returned a main effect for group, $F(1,27) = 12.704$, $p = .001$, $\eta_p^2 = .320$, which makes clear that the CF group made more attempts per question than the NOCF group in all micro sessions (see Figure 3.4). The additional attempts per question made by the CF group appear to be a result of the practice condition. We see at the same time a group by session effect, $F(1,27) = 12.065$, $p = .002$, $\eta_p^2 = .309$, in combination with a session effect, $F(1,27) = 8.165$, $p = .008$, $\eta_p^2 = .232$. A separate analysis of the two groups returns no effect for session in the NOCF group, $F(1,12) = 1.061$, $p = .323$. For the CF group, we find a significant session effect, $F(1,15) = 13.381$, $p = .002$, $\eta_p^2 = .471$, together with a significant session-time interaction effect, $F(2, 30) = 4.984$, $p = .014$, $\eta_p^2 = .249$. The session effect indicates that the CF group significantly decreased the number of attempts per question from the first session ($M = 1.91$) to the second ($M = 1.65$). Posthoc analysis of the session-time effect (Sidak) suggests significant differences between the earlier micro sessions in session 1 and the later micro sessions in session 2: the initial micro session in session 1 is significantly higher ($M = 1.82$) than the third micro session in session 2 ($M = 1.59$), and the second micro session in session 1 ($M = 2.07$) is significantly higher than the second and third micro sessions in session 2 (respectively, $M = 1.64$; $M = 1.59$).

These effects indicate that while no change occurred for the NOCF group, the CF group significantly decreased the number of attempts per question from the first session to the second. The findings for attempts per question above suggest that the CF group expended more effort during practice, but also that they required fewer attempts per question in the second session.

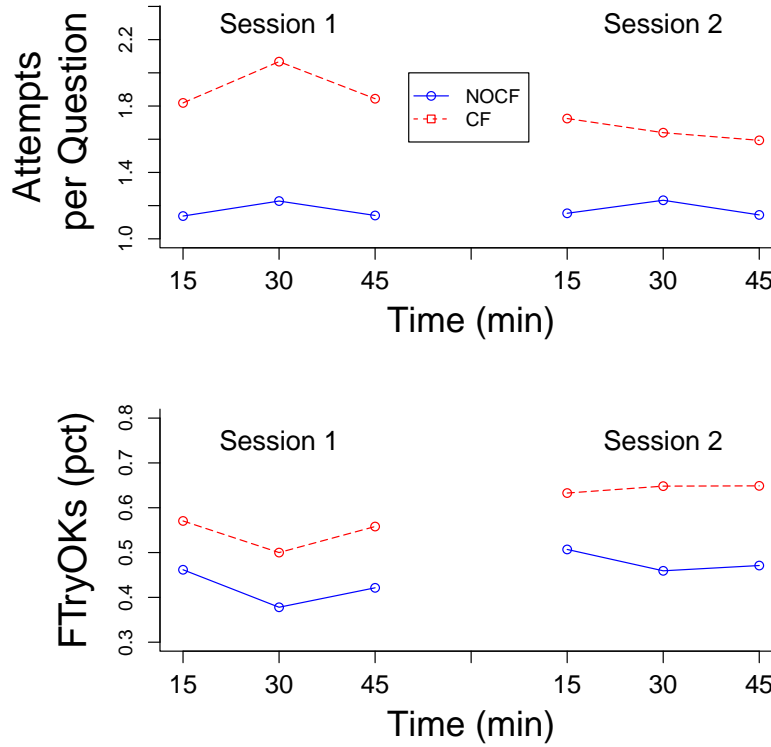


Figure 3.4: Attempts per question and proportion of FirstTryOKs for the NOCF and CF groups for the first, second and third 15 minutes of practice in sessions 1 and 2.

A second point to consider is whether or not learners' performance improved during the course of practice (see Figure 3.4). Analysis of the proportion of FirstTryOKs produced a significant main effect for session, $F(1, 27) = 9.473$, $p = .005$, $\eta_p^2 = .260$, while the remaining main effects for time, $F(2, 54) = 1.763$, $p = .181$, $\eta_p^2 = .061$, and group, $F(1, 27) = 2.853$, $p = .103$, $\eta_p^2 = .096$ were non-significant. No significant interaction effects were observed. The session effect and lack of group interaction effects indicate that both groups improved from session 1 ($M = .487$) to session 2 ($M = .563$).

From this analysis of the log data we can see that participants in the CF group, who were required to correct their word order errors before advancing to the next question, made many more attempts per question. The question answering behaviour of the NOCF group, who tended to attempt each question

only once, did not change over practice, while the CF group reduced the number of attempts they made per question in the second session. Regarding in-practice performance, the number of questions answered correctly on first try increased for both the NOCF and CF groups.

3.5.3 Relationships between practice logs and affective variables

An analysis of the practice measures and their correlations with the affective variables at the global level (see Table 3.1) returned no significant correlations for enjoyment or willingness to practice for either group. For self-efficacy, the analysis showed differences between the groups. For the NOCF group ($N = 13$), the correlations between attempts per question and FirstTryOKs are non-significant. For the CF group ($N = 16$), significant negative correlations were found between attempts per question and self-efficacy at the global level ($r = -.572$, $p = .021$), also present at the session level ($M = -.469$) and in all but one micro session ($M = -.371$), and between self-efficacy and FirstTryOKs ($r = .562$, $p = .024$), present also at the session ($M = .477$) and micro sessions levels ($M = .422$)⁴. Unsurprisingly, in the CF condition these two measures of behaviour are highly correlated ($r = -.956$, $p < .01$), as the CF system automatically advanced learners who answered correctly on first try to the next question. Therefore, these two findings are best treated as one result indicating that in the CF condition there is a relationship between practice performance and self-efficacy. Concerning the apparent difference between self-efficacy and performance relationships in the two groups, it appears that the lack of significant correlations between attempts per question and self-efficacy in the NOCF group is likely due to the lack of variation in the attempts per question (see Table 3.1). However, the lack of relationship between FirstTryOKs and self-efficacy for the NOCF group cannot be explained by a lack of variation in the FirstTryOK data, as both groups had similar standard deviations (see Table 3.1). This difference in relationships between the two groups seems to indicate that the link between learners' performance and their self-efficacy is influenced by CF.

3.6 Discussion

The current study adds to existing research by presenting learners' affective states at the time of practice, obtained via an in-practice questionnaire instrument, and an analysis of how these changed over time. Below, we draw on the results from the in-practice questionnaire and from the practice logs to answer our two research questions.

⁴Values reported for session and micro session levels are the mean of the individual data sets for that level.

Table 3.1: Standard deviation for the two practice measures together with a correlational analysis (Pearson, two-tailed) of the relationship between the measures and affective variables, for the NOCF ($N = 13$) and CF ($N = 16$) groups. The symbol ‘*’ marks relationships which had a significant correlation ($p < .05$).

	<i>SD</i>		Enjoyment		Willingness to Practice		Self-efficacy	
	NOCF	CF	NOCF	CF	NOCF	CF	NOCF	CF
Practice measure								
Attempts / question	0.05 ⁵	0.50	-.07	-.13	-.11	.25	-.40	-.57*
FirstTryOKs	0.19	0.18	.12	.13	.12	-.30	.26	.56*

3.6.1 RQ1: Are there differences in affective states for learners who practice without CF compared to those who practice and receive immediate CF?

The analysis of the in-practice questionnaire data suggests that learner affect in the two practice conditions was comparable. Enjoyment improved in both groups from the first micro session to the second and stayed high throughout the experiment. Willingness to practice fell in both groups within a session, but returned to a high level at the start of the second session and was significantly higher in the second session. Self-efficacy for both groups appears to have increased within the sessions and also from session 1 to session 2. To summarise, CF as provided in this study did not seem to have a general impact on learner enjoyment, willingness to practice or self-efficacy. These are interesting findings which have pedagogical implications we will discuss below.

3.6.2 RQ2: Are there relationships between learner behaviour and learner affect?

An advantage of using a computerised practice environment is that a logging library could be used to automatically record learner and system behaviour which allowed for a detailed look at how practice for the two groups differed. Our analysis indicated that the CF group produced many more attempts per question relative to the NOCF group, that the CF group required fewer attempts per question in the second session while the NOCF group’s behaviour did not change in this respect, and that both groups answered more questions correctly on first try in the second session.

An innovative aspect of this study is that we could combine data from the practice logs and in-practice questionnaires in a correlational analysis to attempt to understand how practice behaviour relates to affective state. With regard to our second research question, we find that only in the

case of the CF practice condition is there a significant relationship between performance-related behaviours and self-efficacy. The same analysis for the NOCF condition produced a similar pattern, but with weaker correlations which did not reach significance. Although learners in the CF condition performed differently, we did not find evidence for a relationship between practice behaviour and enjoyment, nor with willingness to practice, suggesting that these affective dimensions were unrelated to the amount of feedback they received. Summarising, we find that, in the present learning environment, CF did not impact enjoyment or willingness to practice, though it did contribute to a stronger link between performance and self-efficacy. It might be that feedback helps learners gauge their performance during practice. The findings seem to suggest the possibility that, without feedback, learners may improve without recognising their improvement and it may be that CF is then important for building learners' self-efficacy.

In considering our findings, at least two limitations of this study have to be mentioned. First, the feedback was limited to prompt feedback which guided learners towards the correct answer, but did not provide meta-linguistic information on rules for producing correct word order. Cornillie and Desmet (2013) found that learners judged explicit CF with meta-linguistic explanation to be the most useful of the three adopted. Thus, future research could investigate whether or not meta-linguistic information provided in an oral practice context can provide stronger support (in terms of self-efficacy) without negatively impacting enjoyment or willingness to use the system. A second limitation was the relatively small number of participants.

Future investigations employing larger samples and additional analyses would be welcome. For instance, our analyses showed a relationship between practice behaviour and affect for the CF group, but not for the NOCF group. Future research employing video-documented stimulated recall (e.g. MacIntyre & Legatto, 2011) could provide a more explorative analysis of affect and its relationship with practice behaviour.

3.7 Pedagogical implications

From the studies investigating CF and affect, results indicate that learners often perceive error correction somewhat negatively (Kartchava, 2012; Loewen et al., 2009). Our main conclusion here is that CF did not result in negative experiences. CF seems to make learners aware of their performance, which was reflected in their reports of self-efficacy, but did not affect their enjoyment and willingness to practice, suggesting that concerns regarding the negative effects of CF on learner affect (Krashen, 1982; Truscott, 1999) may not apply to the learning environment used in the present study, where the learner is encouraged to experiment.

Apparently, the context created in the present study, an ASR-based CALL system that allows learners to practice L2 speech production and provides

automatic, immediate CF, constitutes a sufficiently safe environment where CF did not evoke a negative experience for L2 learners. This is a finding that might have important implications for future teaching practices. Oral skills are notorious because they require intensive practice that cannot be offered sufficiently in the language classroom and because CF provided in classroom situations might discourage learners. If ASR-based CALL systems turn out to constitute safe and enjoyable learning environments for oral skills with beneficial CF, their use could be stimulated in teaching practice. For instance teachers could decide to delegate intensive, repetitive practice of oral skills to such systems while devoting classroom time to activities that benefit more from the presence of a teacher such as conversational skills.

Finally, we conclude the study on a positive note by observing that, considering grammar practice tends to be viewed as boring but necessary by learners (Jean & Simard, 2011; Loewen et al., 2009), it is an encouraging finding that the computerised oral grammar practice in this experiment appears to have been experienced positively. Since previous studies have also linked this ASR-enabled system (Penning de Vries et al., 2014b, in press) and other environments offering autonomous grammar practice (e.g. Heift, 2003) to developments in L2 proficiency, it may be that emerging CALL systems can provide an effective alternative to traditional grammar practice.

CHAPTER 4

Computerised oral L2 grammar self-study: affective
experience, corrective feedback, initial proficiency and
learning gain

This chapter is to be submitted as

Bodnar, S., Cucchiarini, C., Penning de Vries, B.W.F., Strik, H., & van Hout, R.W.N.M. Computerised oral L2 grammar self-study: affective experience, corrective feedback, initial proficiency and learning gain (to *CALICO Journal*)

Acknowledgements

This chapter has benefitted significantly from the comments of two anonymous reviewers. This work is part of the research program Feedback and the Acquisition of Syntax in Oral Proficiency (FASOP) NWO-Dossier Number 360-75-010, which is funded by the Netherlands Organisation for Scientific Research (NWO).

Abstract

Learning environments that provide positive experiences are held to be linked to positive learning outcomes, but few studies have looked at how corrective feedback, a desired yet often debated characteristic of practice, influences learners' affective experiences in computerised language learning environments. Meanwhile, developments in Automatic Speech Recognition (ASR) technology and research into its application in computer assisted language learning (CALL) have produced systems that can provide immediate feedback on learner speech in stress-free environments. Drawing from recent situated approaches to investigating L2 motivation, the present research examines affective learner experiences in self-study oral CALL grammar practice that provides automatic feedback. Findings from our experiments suggest that feedback provided in an ASR-based CALL system does not have a negative impact on learners' enjoyment, willingness to practice, or self-efficacy, but that individual differences such as initial proficiency and learning gain play a role and that feedback is an aspect of practice important to many learners.

Keywords: L2 speaking, grammar, corrective feedback, learner affect, speech recognition

4.1 Introduction

Learning environments that provide positive experiences are thought to contribute to positive learning outcomes (Gillet, Vallerand, Lafrenière, & Bureau, 2012; Pekrun & Linnenbrink-Garcia, 2014). Work elaborating our understanding of the relationships between second language (L2) practice environments, learners' affective experiences and learning outcomes is needed (Egbert, 2003; Felix, 2008). Fortunately, this topic appears to be gaining interest in computer assisted language learning (CALL) research (e.g. Aerts, Colpaert, & Oberhofer, 2010).

One characteristic of practice that has the potential to impact the learning experience is corrective feedback (CF), which can be defined as information that “draws the learners’ attention to the errors they have made” (Chaudron, 1977, p. 584). CF has been implicated in issues related to L2 learning and learner affect (i.e. emotions or moods) (Agudo, 2013; Krashen, 1982; Truscott, 1999) and to mismatch between teacher and learner beliefs (e.g. Schulz, 2001). Learners frequently report wanting to receive feedback on their errors and a need to receive more (e.g. Cathcart & Olsen, 1976; Schulz, 2001). Possible reasons are classroom constraints (Lasagabaster & Sierra, 2005) or teachers’ hesitation to provide feedback for fear of harming learner affect (Vásquez & Harvey, 2010). The latter is particularly relevant with respect to speaking activities, as these are known to cause anxiety in L2 learners (MacIntyre & Gregersen, 2012; Young, 1990).

Developments in automatic speech recognition technology (ASR) and its application in CALL have produced systems that can provide immediate CF on learners’ spoken utterances in environments that are less likely to provoke anxiety and frustration in case of negative feedback (e.g. Johnson & Valente, 2009; Morton & Jack, 2010; Neri, Mich, Gerosa, & Giuliani, 2008; Penning de Vries, Cucchiari, Bodnar, Strik, & van Hout, 2014b; Wang, Waple, & Kawahara, 2009). So far, little research has investigated learners’ affective experiences in ASR-CALL (Bodnar, Penning de Vries, Cucchiari, Strik, & van Hout, 2013), and the link between CF and affect in L2 learning environments (Ellis, 2010).

In our research on the complex relationship between the development of L2 oral skills and the impact of CF, we have developed an ASR-based system that allows us to tightly control L2 input, L2 output and CF and to monitor multiple features of the learning process. Previous work showed that practice with the system is pedagogically beneficial (Penning de Vries, Cucchiari, Bodnar, Strik, & van Hout, 2014a; Penning de Vries et al., 2014b). In a study of in-practice affect (see Chapter 3), we investigated learners’ enjoyment, willingness to practice and self-efficacy on a micro-scale by observing how they changed within 45-minute practice sessions and by searching for relationships between affective states and logs of user-system interactions. We found that learners who practiced with feedback produced more utterances as a result of being forced to correct their errors, but that this did not have a negative impact

on their affective practice experience, suggesting that CF can be provided safely in this type of computerised learning environment.

The present study steps back from the micro-perspective taken in Chapter 3 to perform a more comprehensive evaluation of the same computerised self-study environments (i.e. practice with and without automatic CF) that includes analyses of in-practice affect, a post-practice evaluation, initial proficiency and learning gain, and the relationships between these components.

4.2 Research background

Gardner (1985b) was one of the first to investigate how affective variables, or “the emotionally relevant characteristics of the individual that influence how she/he will respond to any situation” (p. 1) influence L2 acquisition. Gardner’s research included a number of affective factors, but is best known for studying the relationship between learners’ attitudes towards a native speaker population and long-term language achievement (Dörnyei & Ushioda, 2011).

Recently, SLA researchers have expanded their study of affective variables to include self-confidence (Clément, 1986), willingness to communicate (WTC) with native speakers (MacIntyre, Burns, & Jessome, 2011), and foreign language anxiety and enjoyment (Dewaele and MacIntyre, 2014). Traditionally, these factors have fallen under the rubric of affective individual differences. The domain covered by the term ‘affect’ in SLA (e.g. Arnold & Brown, 1999) is broader than that in emotion or educational psychology research where it is used to refer to enjoyment, boredom and other “emotions and moods” (Pekrun & Linnenbrink-Garcia, 2014, p. 3).

Affective characteristics can be distinguished based on their time scale: attitudes are considered to be relatively stable and context-independent (Shuman & Scherer, 2014). The term ‘state’ has been used to refer to less stable affective characteristics that fluctuate with time and context (e.g. Tremblay, Goldberg, & Gardner, 1995). Examples of affective states in SLA literature include emotional states such as experiencing fun, frustration, and embarrassment (Agudo, 2013; Garrett & Young, 2009), as well as more complex, multi-componential states such as “feelings of self-confidence, feeling willing to communicate, or feeling anxious” (MacIntyre & Gregersen, 2012, p. 103), or interest (Ainley & Hidi, 2014).

The present study employs the adjective affective to mean “broadly relating to emotions” (e.g. ‘affective variables’) and the noun affect in the more narrow sense that MacIntyre and Gregersen (2012) use it, to refer to basic emotional states such as pleasure and other transient complex affective states such as WTC¹.

Although for many years the more stable affective characteristics have

¹While this use of ‘affect’ is in-line with previous work in the area, some researchers are now differentiating between affect, motivation and cognition (e.g. Dörnyei, 2009b, see also Dörnyei & Ushioda, 2011).

been addressed in SLA, recently there has been increasing interest in situated perspectives on motivation, which focus on language learning in specific situations with shorter time scales. Ushioda (2009b) stresses the importance of considering learners as they interact with their environment and suggests that learners' affective experiences in a particular context be studied in conjunction with objective records of the events that take place (Ushioda & Dörnyei, 2012). In our view, the usefulness of situated perspectives on L2 motivation can be extended to other affective states and the present study adopts this perspective.

Previous studies have investigated affective variables in computerised L2 practice, but few have adopted a situated perspective (Bodnar, Cucchiarini, Strik, & van Hout, 2016). Jauregi, de Graaff, van den Bergh, and Kriz (2012) investigated learners' subjective experiences in synchronous video conferencing with native speakers taking place in the context of a traditional language course for learners of Dutch. The study tracked changes in learners' WTC, attitudes towards the course and attitudes towards the L2 culture. Questionnaire results show that beginning learners experienced positive changes in linguistic confidence and a reduction in language anxiety during the weeks containing the video conferencing sessions, while the advanced learners did not receive these benefits.

A number of constructs seem particularly interesting in this respect. A first one is whether or not learners enjoy practice with a specific system, i.e. whether they think it is 'fun' or 'nice' (Dewaele and MacIntyre, 2014; see also Ainley & Hidi, 2014) to practice. Enjoyment is thought to foster learning (Dewaele and MacIntyre, 2014), and while CF has been mentioned in conjunction with negative affective experiences in classroom research, this has rarely been explored in CALL environments. A second relevant construct is a learner's willingness to practice with a system. This component is related to state motivation capturing "an individual's motivational condition at a particular point in time" (Tremblay et al., 1995, p.358), dynamic approach-avoidance motivation (P. MacIntyre & Serroul, 2015) and WTC (MacIntyre & Legatto, 2011) adapted to the context of oral L2 self-study. A third construct is self-efficacy (Bandura, 1978), which refers to the beliefs learners have about their ability to accomplish specific tasks (Bandura, 1978). Studies have shown that perceived self-competence during language learning is a significant factor influencing learner motivation (Busse & Walter, 2013; Ushioda, 1994). Tracking self-efficacy can provide information on whether or not learners' feel competent during practice and how this changes over time. More information on these constructs is available in Chapter 3.

CF has received a great deal of interest in recent years (Lyster et al., 2013; Sheen & Lyster, 2010), and has been studied also in CALL environments (Cornillie, Clarebout, & Desmet, 2012; Cornillie & Desmet, 2013; Heift, 2013; Sagarra, 2007; Sagarra & Abbuhl, 2013). So far, little research has addressed the interaction of CF, affective states and learning outcomes in an ASR-based CALL context, in which learners can practice oral skills and receive online feedback from the system in a less threatening environment. The impact

of CF on self-efficacy was the topic of a study by Chan and Lam (2010) that investigated how self-efficacy is prejudiced by different types of feedback in L2 vocabulary assessments. A computerised vocabulary test was used that provided two different types of feedback on test performance: formative feedback informed learners about the questions they answered incorrectly, while summative feedback notified learners of how well they performed. Results suggested that the self-efficacy of both groups decreased from pre-test to post-test, but that it decreased less for the group receiving formative feedback, supposedly because this provided learners with a means to improve, which is lacking in summative feedback.

Cornillie et al. (2012) evaluated the perceived usefulness of three types of progressively less explicit CF provided during written conversation simulations with computer-controlled characters in a virtual 3D world. Analyses indicated that participants found CF useful, that explicit CF was the most useful (despite being less game-like), and that people who appreciated explicit CF tended to be more intrinsically motivated, have a more positive game experience and perceive themselves to have completed the game more competently.

Cornillie and Desmet (2013) looked at how two important gaming features, fantasy and vividness, can be incorporated into CF in form-focused practice to provide a positive failure experience that maintains motivation. Preliminary analyses suggested that fantasy and vividness could be associated with increased immersion in the game, but interviews revealed that no participant preferred the vivid feedback, with some saying it was distracting or frustrating.

Summarising, the literature on feedback and affect in computerised self-study environments is limited, but the limited findings obtained so far and those emerging from adjacent domains (Aghaei Pour, Hussain, AlZoubi, D'Mello, & Calvo, 2010; Narciss, 2004) provide evidence for a link between feedback and perceived self-competence or self-efficacy, for positive and negative emotional responses to different kinds of feedback, and for a relationship between CF, affective states, and learning outcomes.

4.3 The present study

The present study addresses some of the particular issues in the SLA literature concerning students' desire to receive feedback on their oral productions while simultaneously tending to have negative experiences when receiving feedback. We hypothesized that a CALL-ASR system providing oral self-study grammar practice would allow learners to receive feedback on their errors without the negative experiences associated with feedback in the classroom, and that feedback would contribute to a positive experience on the grounds that it provides learners with an opportunity to improve in a non-threatening environment. Our research questions were the following:

1. Does the oral grammar practice provided by the system result in learning gains and positive affective experiences?

2. What impact does automatic CF on oral productions have on learning gains and affective experiences?
3. How do learners' initial proficiency levels and learning gains relate to affective experiences?

This study proceeds in two parts: Part one investigates the effectiveness of practicing in a CALL-ASR environment in the presence or absence of immediate CF: an analysis of learning gains is followed by an examination of learners' self-reported enjoyment, willingness to practice and self-efficacy levels and their relationships to a post-evaluation of the environment. Part two focuses on the learners by investigating the relationship between initial proficiency, learning gain and affective experiences.

4.4 Materials and methods

Data for the present study was collected in three comparable experiments. Materials used in the experiments are presented first, followed by descriptions of the participants, experimental procedures and the steps used in preparing the data for analysis.

4.4.1 Target L2 feature

The experiments employed a practice environment designed to provide L2 learners of Dutch with self-study oral grammar practice focusing on one specific aspect of Dutch word order, verb second (i.e. inversion of the subject and verb position when a phrase begins with another phrase constituent), which is known to be difficult for learners of Dutch (Jordens, 1988). Because deviations from this rule rarely cause communication breakdown, learners may remain unaware of their errors in spoken situations.

We designed exercises consisting of pairs of video clips and oral quizzes (see Figure 4.1). Learners first watch part of a story in the video clip and answer questions about the video by speaking their answers aloud. The oral quiz component is a word block ordering exercise: learners first read a question and then attempt to produce a grammatically correct utterance from the word blocks on the screen. More information on the exercise is available in (Penning de Vries et al., 2014b).

4.4.2 Automatic CF

Providing automatic CF in oral L2 practice requires the use of ASR to convert oral productions to text so they can be automatically evaluated for grammatical correctness, which is particularly challenging in the case of L2 learners (van Doremalen, Strik, & Cucchiaroni, 2010). In the present system, a feedback



Figure 4.1: A screenshot of the Dutch word order exercise. The system presents learners with a video clip, after which learners answer questions about the clip by speaking Dutch. Learners read the question (1), mentally arrange the word blocks (2) into a correct order, and speak their answer aloud after pressing the record button (3).

module responds to recorded answers with CF that provides information on the correctness of the word block orderings.

If the answer is correct learners receive confirmatory feedback (Figure 4.2, left), otherwise one of two types of feedback is provided, depending on the system configuration. The first type is modeled after elicitation feedback (Lyster & Ranta, 1997). It signals the incorrect answer and prompts the learner to try again (Figure 4.2, centre). From their second mistake onwards, they also receive a hint whereby the next word block is automatically placed in the answer area. The second type of feedback is modeled after an explicit recast (Erlam & Loewen, 2010). The system displays a message notifying the learner of an error and displays the correct ordering of the blocks (Figure 4.2, right).



Figure 4.2: A screenshot of feedback in the system. A feedback module notifies the learner that their answer is either correct (left), or incorrect. Two types of CF are available: prompt (middle) or recast (right).

4.4.3 In-practice questionnaire

To tap into learners' affective experiences during practice, we developed a short questionnaire designed to poll their affective states at regular intervals during practice (e.g. every 15 minutes). With such a design, it is important to ensure the questionnaire is minimally interruptive (Dela Rosa & Eskenazi, 2011). The questionnaire was therefore limited to three constructs, enjoyment, willingness to practice and self-efficacy, which were expected to be sensitive to the learning environment and CF. In line with situated perspectives on L2 motivation (Ushioda & Dörnyei, 2012) highlighting the importance of taking the learning context into account, an effort was made to formulate questionnaire items specific to the CALL environment (cf. Dela Rosa & Eskenazi, 2011). Three five-point semantic differential items were used to survey in-practice affect (see Figure 4.3):

- Enjoyment: *Oefenen met GREET is ... (Helemaal niet leuk - Heel leuk)*
Practising with GREET is ... (Not nice at all - Very nice)²
- Willingness to practice: *Op dit moment wil ik ... (Iets anders doen - Verder oefenen met GREET)*
At this moment I want to ... (Do something else - Continue practising with GREET)
- Self-efficacy: *De volgende vragen zullen ... zijn voor mij (Moelijk [sic] - Makkelijk)*
The next questions will be ... for me (Difficult - Easy)

4.4.4 L2 proficiency tests

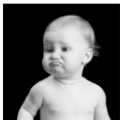
To measure changes in learners' V2 accuracy, we developed a computerised spoken Discourse Completion Test (DCT) (see Figure 4.4) consisting of 32 items, with 16 target items testing specifically for V2 and 16 filler items (Penning de Vries et al., 2014a, 2014b). Two different versions of the test, (A and B) containing linguistically equivalent, but not identical items were randomly assigned to participants in the pre-test and the post-test.

4.4.5 Post-practice questionnaire

To examine learners' overall impressions of practice with the system, we developed a second questionnaire to be completed on the last day of the experiment, after the proficiency tests. Since in this case we did not have to worry about interrupting practicing learners, more items probing learners' evaluation of different aspects of the practice were included. Two versions

²The first time this item was presented, the slightly different wording of *Ik denk dat het oefenen met GREET ... zal zijn (Helemaal niet leuk - Heel leuk)*. I think that the exercise with GREET will be ... to reflect that the user had not yet practiced with the system


Oefenen met GREET is _____ [English]




Helemaal niet leuk [English]

☐ ☐ ☐ ☒ ☐

Heel leuk [English]




Op dit moment wil ik _____ [English]




Iets anders doen [English]

☐ ☐ ☐ ☒ ☐

Verder oefenen met GREET [English]



De volgende vragen zullen _____ zijn voor mij. [English]



Moelijk [English]

☐ ☒ ☐ ☐ ☐

Makkelijk [English]




Figure 4.3: A screenshot showing the in-practice questionnaire, with items for enjoyment, willingness to practice and self-efficacy (see text for translations).

were used: A CF version contained questions about learners' experience of receiving feedback from the system, while a NOCF version contained questions on the importance learners attach to feedback. The two questionnaire versions shared seven items which we used for comparisons between the NOCF and CF conditions:

1. *Ik vond het . . . om met GREET te werken (Saai-Leuk)*
I thought it was . . . to work with GREET (boring-fun)
2. *Oefenen met GREET is een . . . manier om mijn Nederlands te verbeteren (Slechte-Goede)*
Practicing with GREET is a . . . way to improve my Dutch (bad-good)
3. *Mijn Nederlands is . . . geworden door te oefenen met GREET (Slechter-Beter)*
My Dutch has become . . . as a result of practicing with GREET (worse-better)
4. *De videos zijn . . . (Saai-Leuk)*
The videos were . . . (boring-fun)

FASOP spreektoets

Vraag 1
1 / 32

"De band van ... "



fiets + lek

Neem je antwoord op:

RECORD

resterende tijd:

Volgende vraag

Figure 4.4: A screenshot of the DCT. Each item consists of an unfinished Dutch sentence (left) together with a picture and hint to add context (right). Learners complete the item by formulating and then recording an answer before the time bar reaches its limit (bottom).

5. *De vragen waren . . . (Saai-Leuk)*
The questions were . . . (boring-fun)
6. *De videos waren . . . (Moeilijk-Makkelijk)*
The videos are . . . to understand (difficult-easy)
7. *De vragen waren . . . (Moeilijk-Makkelijk)*
The questions were . . . (difficult-easy)

4.4.6 Participants

Adult learners of Dutch with proficiency levels A1 and A2 of the Common European Framework (CEF) studying at language institutes near the university were recruited. These participants had a variety of L1s and cultural backgrounds. In total 31 participants were recruited for experiment 1, 30 for experiment 2, and 27 for experiment 3. In experiment 2 only, a group of German participants following an intensive Dutch course was included ($N = 10$) because of their low level of Dutch and prior experience learning English. Although German native speakers might be expected to be proficient with V2, some work shows that

learners of L3 German with prior exposure to English can have difficulty with V2 (Bohnacker, 2006).

4.4.7 Procedures

In experiment 1, learners visited the lab twice. In the first session, they completed a background questionnaire, two proficiency tests and a 45-minute practice session. Learners were randomly assigned to a CF-enabled condition and a CF-disabled (NOCF) condition and practiced for 45 minutes. Once before practice and every 15 minutes thereafter, the system paused the exercise and required the learners to fill out the in-practice questionnaire. In the second session, participants first completed practice, again filling out in-practice questionnaires, and then completed the proficiency tests followed by the post-practice questionnaire.

The experimental design for experiment 2 was identical to that of experiment 1, with one exception: learners were randomly assigned to either a CF condition providing prompt feedback (CF-P), or a CF-R condition providing recast feedback. All other aspects of the experiment were identical to experiment 1.

Experiment 3 was aimed at studying CF over a longer period of time. A total of seven 30-minute sessions were scheduled. Learners completed the background questionnaire and proficiency tests in the first session and the post-tests and post-practice questionnaire in the final session.

4.4.8 Data analysis procedures

Questionnaire data was extracted from the system database and learner proficiency was obtained by manually scoring the pre-test and post-test recorded utterances for word order correctness. Learning gain was calculated by subtracting the pre-test score from the post-test score. Together, the three experiments yielded data for 88 participants. This data was reviewed to remove participants who had technical difficulties (2), missed practice sessions (7), or had high pre-test scores indicating they had little to learn from the experiment (23)³. Excluding these participants left a total of 56 learners for the study.

Due to the low number of participants in each experiment group after filtering, it was decided to pool the data from the three experiments. To enable the pooling, we decided to compare the in-practice questionnaire data for the first and last practice sessions of each experiment. This was done because the third experiment included a different number of practice sessions and in-practice questionnaires compared to the first two experiments. The mean of the in-practice questionnaire data recorded during the first and last practice sessions provided two data points for each participant.

³High pre-test scores indicate mastery of V2 inversion and complicate the analysis with ceiling effects. Inspecting background data for the 23 learners revealed that most were indeed native speakers of German (10) or had learned it as an L2 (7).

4.5 Results Part one: Affective experiences in CALL-ASR grammar practice and learning gains

4.5.1 Learning gains

A repeated measures ANOVA analysis of proficiency with factors session (within-subjects factor) and practice condition (CF or not, between-subjects factor) returns two effects (see Figure 4.5). First, there is a significant main effect of session, $F(1, 54) = 38.74$, $p = .000$, $\eta_p^2 = .418$, indicating learning gain: learners' scores increased from the first session ($M = 5.41$, $SD = 4.19$) to the last session ($M = 8.21$, $SD = 5.11$). Second, a significant main effect for practice condition, $F(1, 54) = 4.127$, $p = .047$, $\eta_p^2 = .071$, reveals that the NOCF group ($M = 4.0$, $SD = 4.1$) is significantly less proficient than the CF group ($M = 6.14$, $SD = 4.11$) at pre-test and at post-test (respectively, $M = 6.37$, $SD = 5.64$; $M = 9.16$, $SD = 4.62$). The interaction of session and practice condition is not significant, $F(1, 54) = .577$, $p = .451$, $\eta_p^2 = .011$, showing that the difference between the groups does not change over the experiment. Summarising, it seems learners practising in the NOCF condition were less proficient compared to learners in the CF condition, and that both groups improve from first to last session⁴. As learners were assigned randomly to the practice conditions, no difference in initial proficiency was expected between the conditions. The difference, being just significant with a low η_p^2 value, seems to be the outcome of sample fluctuation, the exclusion of high proficient learners, and drawing learners from different experiments.

4.5.2 Affective experiences: in-practice questionnaire data

A correlation analysis (Pearson) of the relationships between the in-practice questionnaire items reveals moderate correlations (see Table 4.1). The positive relationship between enjoyment and willingness to practice ($p < .01$) suggests people who enjoy practice tend to be willing to use the system. Similarly, the positive relationship between enjoyment and self-efficacy ($p < .01$) suggests learners who enjoy practice tend to feel competent in the exercises. Learners' willingness to practice with the system appears unrelated to their self-efficacy.

In line with these moderate correlations, we decided to treat the items as separate constructs. Calculation of the overall means for enjoyment, willingness to practice and self-efficacy for the learners ($N = 56$) suggests generally positive results (see Table 4.2). Enjoyment scores high ($M = 4.03$, $SD = 0.853$), indicating that learners enjoyed practice; learners were generally willing to practice with the system ($M = 3.79$, $SD = 1.05$); self-efficacy, being closer

⁴This pattern applies to all three experiments. We reported in earlier articles an interaction effect (Penning de Vries, Cucchiari, Bodnar, Strik, & van Hout, in press)

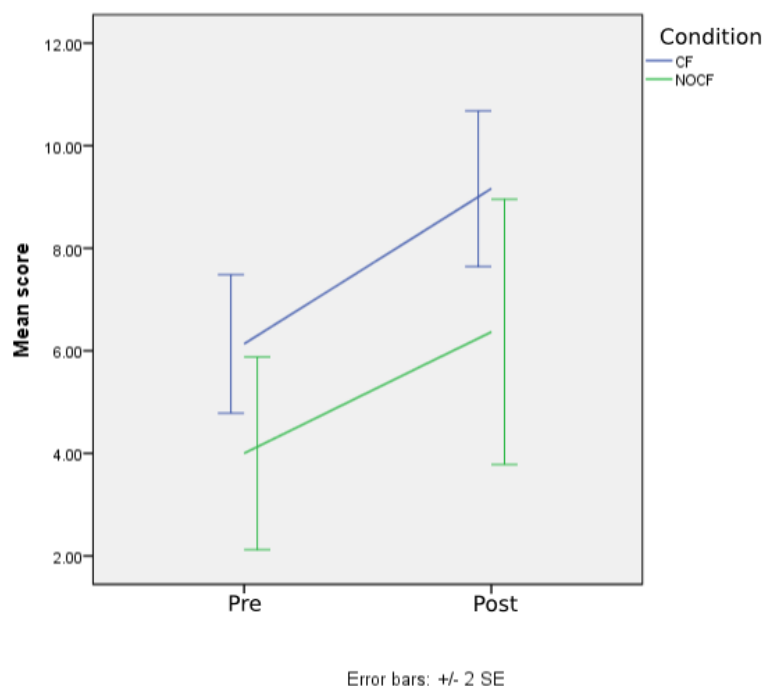


Figure 4.5: DCT proficiency scores for the NOCF (green) and CF conditions (blue) before (Pre) and after (Post) using the system. Maximum possible score was 16.

to neutral ($M = 3.2$, $SD = 0.963$), suggests that learners tended to predict that questions would be neither too hard, nor too easy.

A repeated measures ANOVA with factors session and practice condition revealed no significant main or interaction effects for the three items (see Table 4.2). The lack of any effects whatsoever suggests that these affective dimensions of practice were similar for both conditions and in both sessions.

4.5.3 Learning experiences: post-practice questionnaire

A reliability analysis of the post-practice questionnaire indicates moderate reliability (Cronbach's $\alpha = .738$), opening up the possibility of investigating the items for underlying subscales. A principal component analysis with varimax rotation produced two factors, shown in Table 4.3. Based on the loadings of the items, we defined two separate indices:

1. Attractiveness (items 1, 2, 3, 4 and 5) captures learners' overall evaluation

Table 4.1: Overall means, standard deviations and correlations for enjoyment, willingness to practice and self-efficacy. ** Correlation is significant at the 0.01 level (two-tailed).

	<i>M</i>	<i>SD</i>	Enjoyment	Willingness to practice	Self-efficacy
Enjoyment	4.03	0.85	1	.67**	.37**
Willingness to practice	3.79	1.05	—	1	.18
Self-efficacy	3.20	0.96	—	—	1

Table 4.2: Repeated measures ANOVA analysis with factors practice condition and session of the in-practice questionnaire data finds no significant effects.

Effect	Enjoyment	Willingness to practice	Self-efficacy
Practice	F(1,54) = .008,	F(1,54) = .657,	F(1,54) = .004,
Condition (PC)	$p = .929$	$p = .421$	$p = .950$
Session	F(1,54) = .263,	F(1,54) = 1.359,	F(1,54) = 2.839,
	$p = .610$	$p = .249$	$p = .098$
PC x Session	F(1,54) = .066,	F(1,54) = .032,	F(1,54) = .434,
	$p = .799$	$p = .858$	$p = .513$

of the system;

- Difficulty (items 6 and 7) reflects learners' perceptions regarding practice difficulty.

Calculating scores for the two scales by averaging scores for the items belonging to each component (see Table 4.4) produced overall positive findings: participants find the practice to be attractive ($M = 3.90$, $SD = 0.74$), and the exercises were moderately easy ($M = 3.79$, $SD = 0.77$) (note that extreme values here would indicate that the questions were either very difficult or very easy).

Next, one-way ANOVAs were used to investigate the effect of practice condition on ratings of attractiveness and difficulty. Results indicate no significant differences between the groups for attractiveness, $F(1,54) = .326$, $p = .570$, $\eta_p^2 = .006$. However, the analysis finds a significant difference for difficulty, $F(1,54) = 5.1$, $p = .028$, $\eta_p^2 = .086$, suggesting that the NOCF group reported more difficulty ($M = 3.47$, $SD = 0.79$) than the CF group ($M = 3.95$, $SD = 0.71$). Because prompt feedback (CF-P) withholds the correct form while the recasts (CF-R) provide it, we also investigated the difficulty between the two types, finding a significant practice condition effect, $F(2,53) = 4.87$, $p = .011$, $\eta_p^2 = .155$. Post-hoc analysis (Sidak) finds that the CF-R condition

Table 4.3: Principal components analysis after varimax rotation and Kaiser normalisation of the post-practice questionnaire. Items 1 to 5 load on factor 1. Items 6 and 7 load on factor 2.

Item	Component	
	1	2
1. I thought it was ...to work with GREET (boring-fun)	.814	.209
2. Practicing with GREET is a ... way to improve my Dutch (bad-good)	.831	-.207
3. My Dutch has ... as a result of practicing with GREET (worsened-improved)	.776	.046
4. The videos were ... (boring-fun)	.625	.262
5. The questions were ... (boring-fun)	.523	.411
6. The videos are ...to understand (difficult-easy)	.069	.796
7. The questions were ... (difficult-easy)	.063	.780

Table 4.4: Means and standard deviations for the post-practice questionnaire indices. Higher difficulty index values indicate less difficulty.

Index	<i>M</i>	<i>SD</i>	Items
Attractiveness	3.90	0.74	1, 2, 3, 4, and 5
Difficulty	3.79	0.77	6 and 7

is the easiest ($M = 4.35$, $SD = 0.474$), significantly different from the NOCF condition ($M = 3.474$, $SD = 0.79$), while the differences between the CF-R and CF-P ($M = 3.8$, $SD = 0.737$) and CF-P and NOCF conditions are too small to be significant.

4.5.4 Relationships between in-practice and post-practice questionnaire data

A correlational analysis (Pearson) of the in-practice and post-practice questionnaire constructs (see Table 4.5) produces significant positive correlations between enjoyment and attractiveness, $r(54) = .403$, $p = 0.002$, and self-efficacy and difficulty, $r(54) = .393$, $p = .003$. These relationships seem plausible and there is also a lack of implausible relationships that would suggest invalid measures. The lack of strong correlations can be explained by the generally high scores for all participants (ceiling effect).

Table 4.5: Correlations between the in-practice and post-practice questionnaire constructs. ** significant at the 0.01 level (two-tailed).

	Attractiveness	Difficulty
Enjoyment	.40**	.03
Willingness to practice	.21	-.11
Self-efficacy	-.04	.39**

4.6 Results Part two: Relating in-practice affective experiences to initial proficiency and learning gains

4.6.1 Initial proficiency and learning gains

We begin by creating categories to differentiate between learners who improve on the proficiency tests from those who do not improve. This procedure requires a threshold, a value reflecting the minimum amount of change we accept as improvement. To minimize the role of chance in determining a learning outcome, we chose a threshold of an increase of three or more points (+3) as an indicator of improvement.

Another learner characteristic which might have influenced learners' affective states is their initial proficiency. Learners included in this study had pre-test scores from 0 to 12 inclusive (see section 4.7). We divided this range equally and labeled learners low proficient if they scored below 6 points, or medium proficient if they scored above 6 and below 13 points. This assignment produced a comparable number of low- and medium-proficient learners (30 and 26 respectively). Figure 4.6 visualises the different learner proficiencies and learning outcomes.

Table 4.6: The distribution of learners across the learning gains and initial proficiency categories. Numbers in parentheses indicate the number of participants in the practice conditions (NOCF, CF).

Proficiency	Improve ($x \geq +3$)	Not improve ($x < +3$)	Total
Low ($x < 6$)	15 (5, 10)	15 (9, 6)	30 (14, 16)
Med ($6 < x < 13$)	15 (4, 11)	11 (1, 10)	26 (5, 21)
Total	30 (9, 21)	26 (10, 16)	56 (19, 37)

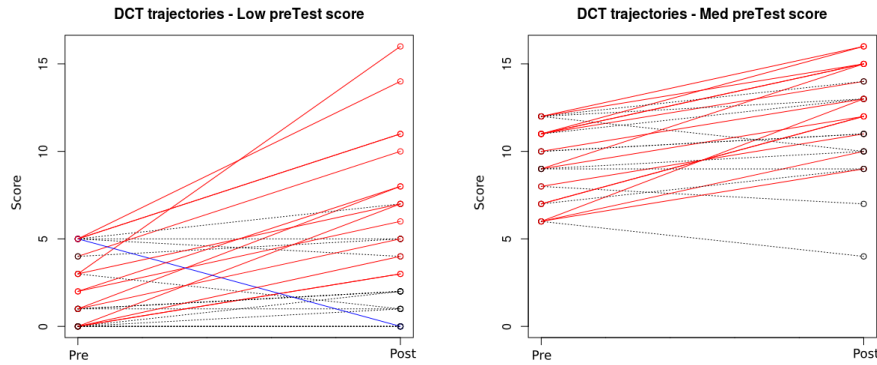


Figure 4.6: Learning trajectories in the experiment. To differentiate between learners who improve (red lines) and those who do not improve, i.e. those with no change or who worsen (marked with black dashed and blue solid lines respectively), we use a threshold of an increase of three. Most people improve (30), though a large number do not improve (26).

4.6.2 In-practice questionnaire: affective experiences

A repeated measures ANOVA was run including a total of four factors: initial proficiency (low vs. medium), learning outcome (improve vs. not improve), practice condition (NOCF or CF) and session (1 or 2). The analysis returned complex effects with multiple interactions. Differing cell sizes (see Table 4.6) prevented a meaningful analysis of a cross of all the factors, although it was clear in this first stage of the analysis that there were differences. We noted that initial proficiency was involved in a number of interaction effects for enjoyment, which had the most complex interactions, and willingness to practice. Splitting the analysis between low and medium proficient learners simplified the outcomes and helped to understand the trends within the groups, as hardly any effects were found between the low proficiency groups, whereas clear differences were found between the medium proficiency groups.

Enjoyment

Analysis of enjoyment for the low proficiency group reveals no significant main effects for session, $F(1, 26) = 2.493$, $p = .126$, $\eta_p^2 = .088$, practice condition, $F(1, 26) = 0.000$, $p = .987$, $\eta_p^2 = .000$, or learning outcome, $F(1, 26) = 0.018$, $p = .894$, $\eta_p^2 = .001$, nor were any significant interaction effects observed.

For the medium proficiency learners, the analysis also did not return any significant main effects, for session, $F(1, 22) = 0.002$, $p = .968$, $\eta_p^2 = .000$, practice condition, $F(1, 22) = .763$, $p = .392$, $\eta_p^2 = .034$, or learning outcome, $F(1, 22) = .462$, $p = .504$, $\eta_p^2 = .021$. However, the analysis revealed a significant

interaction effect between factors session and learning outcome (see Figure 4.7), $F(1, 22) = 5.634$, $p = .027$, $\eta_p^2 = .204$. Post-hoc analysis (Sidak) indicated that, for the group of learners who did not improve, enjoyment decreased significantly from the first session ($M = 4.175$, $SD = 0.780$) to the last ($M = 3.65$, $SD = 1.035$), while for the group of learners who improved, the enjoyment reported in the first session ($M = 4.117$, $SD = 0.831$) remained similar to that reported in the last ($M = 4.361$, $SD = 0.644$). The other interaction effects were not strong enough to be significant.

These findings suggest that all learners enjoyed using the system during the first session, and that most learners continued to enjoy practice in the last session despite the exercises being similar. For medium proficiency learners who did not improve, however, enjoyment decreased significantly in the last session.

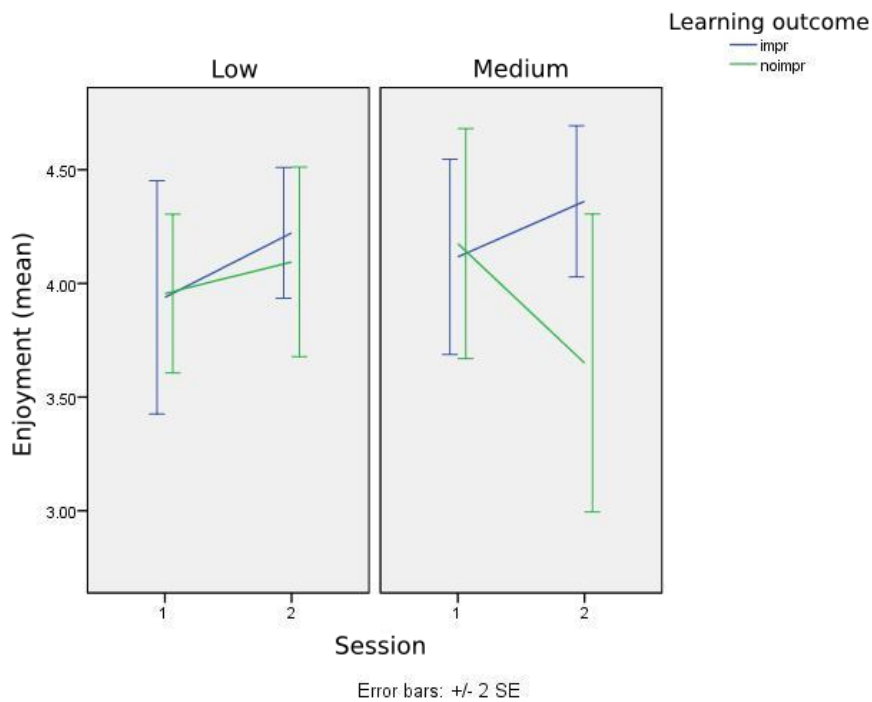


Figure 4.7: Enjoyment as related to initial proficiency and learning gains. Left panel: enjoyment levels for low proficiency groups were comparable in both sessions. Right panel: enjoyment for the medium proficiency group of learners who did not improve during practice decreases significantly, while enjoyment for the group who improved does not change significantly.

Willingness to practice

Analysis of willingness to practice in the low proficiency group reveals no significant main effects: the effect for session was not significant, $F(1, 26) = 2.794$, $p = .107$, $\eta_p^2 = .097$, nor were the effects for practice condition, $F(1, 26) = 0.169$, $p = .685$, $\eta_p^2 = .006$, or learning outcome, $F(1, 26) = 0.19$, $p = .666$, $\eta_p^2 = .007$. No significant interaction effects were observed.

Analysis of willingness to practice for medium proficiency learners reveals a significant effect for learning outcome, $F(1, 22) = 6.581$, $p = .018$, $\eta_p^2 = .230$, indicating that the group who learned ($M = 4.019$, $SD = 0.932$) was significantly more willing to practice than the group who did not learn ($M = 2.988$, $SD = 1.157$; see Figure 4.8). The remaining main effects were non-significant (session, $F(1, 22) = 0.376$, $p = .546$, $\eta_p^2 = .017$; practice condition, $F(1, 22) = 0.802$, $p = .38$, $\eta_p^2 = .035$). No significant interaction effects were observed.

These findings suggest a relationship between the medium proficiency learners' willingness to practice and their learning outcomes, while this relationship is absent for the low proficiency learners.

Self-efficacy

Analysis of self-efficacy for low proficiency learners revealed a significant main effect for session, $F(1, 26) = 5.123$, $p = .032$, $\eta_p^2 = .165$, indicating that self-efficacy increased significantly from the first session ($M = 2.711$, $SD = 0.873$) to the last session ($M = 3.213$, $SD = 0.897$). The factor practice condition was not significant, $F(1, 26) = 0.569$, $p = .457$, $\eta_p^2 = .021$, nor was learning outcome, $F(1, 26) = .543$, $p = .468$, $\eta_p^2 = .020$. No significant interaction effects were observed.

Analysis of self-efficacy for medium proficiency learners revealed that the main effects are not significant, for session, $F(1, 22) = 0.827$, $p = .373$, $\eta_p^2 = .036$, practice condition, $F(1, 22) = 0.250$, $p = .622$, $\eta_p^2 = .011$, and learning outcome, $F(1, 22) = 1.145$, $p = .296$, $\eta_p^2 = .049$. However, a significant interaction effect was observed between session and learning outcome (see Figure 4.9), $F(1, 22) = 6.786$, $p = .016$, $\eta_p^2 = .236$, with post-hoc analysis (Sidak) indicating that, for the group of learners who improved, self-efficacy increased significantly from the first session ($M = 3.3833$, $SD = 0.979$) to the last session ($M = 3.956$, $SD = 0.857$). For learners who did not improve, self-efficacy levels appear to decrease from the first session ($M = 3.500$, $SD = 0.764$) to the last session ($M = 2.950$, $SD = 1.046$), and while post-hoc analysis (Sidak) reveals this difference to be non-significant, it also indicates that the difference between the improvers and non-improvers in the last session is significant⁵, pointing to a divergence between the two groups.

These findings seem straightforward, suggesting that low proficiency learners generally reported higher self-efficacy in the last session whether or not they

⁵Note that this is based on SPSS Estimated Marginal Means

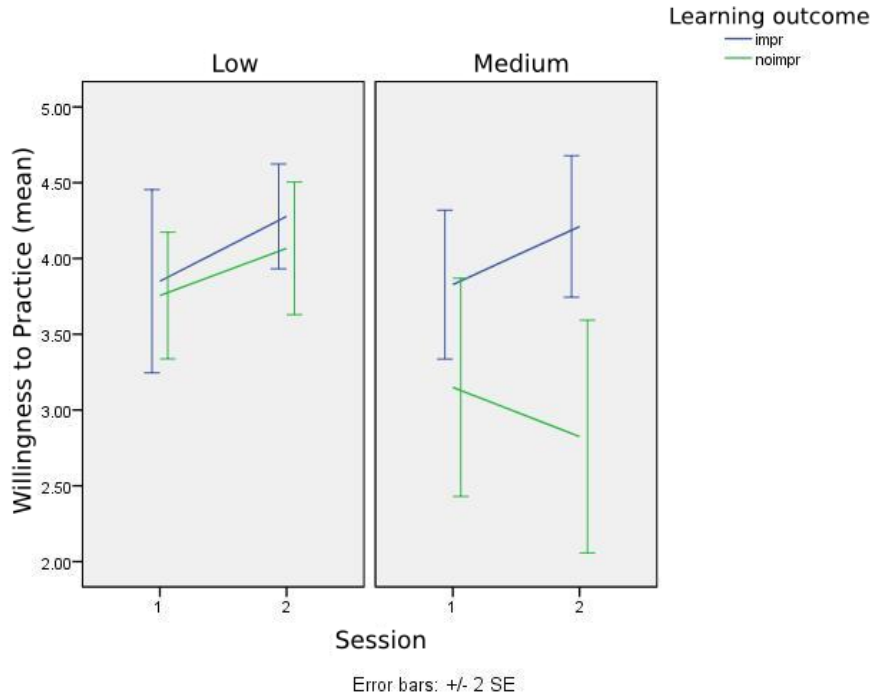


Figure 4.8: Willingness to practice as related to initial proficiency and learning gains. Left panel: willingness to practice levels for the low proficiency groups are comparable in both sessions. Right panel: In the first session, willingness to practice levels for the medium proficiency learners who do not improve are significantly lower than those of learners who improve and this difference remains in the last session.

actually improved, while in the case of the medium proficiency learners, only those who improved reported an increase in self-efficacy.

Post-practice questionnaire

Analysis of attractiveness ratings obtained from the post-practice questionnaire for the low and medium proficiency learners returns main effects that are not significant, but also a significant interaction effect between practice condition and proficiency (see Figure 4.10), $F(1, 48) = 6.05$, $p = .018$, $\eta_p^2 = .112$. Post-hoc testing reveals the low proficiency group practising in the NOCF condition judged practice to be less attractive ($M = 3.61$, $SD = 0.72$) than their counterparts in the CF condition ($M = 4.23$, $SD = 0.53$).

Analysis of the difficulty index returns a significant main effect for

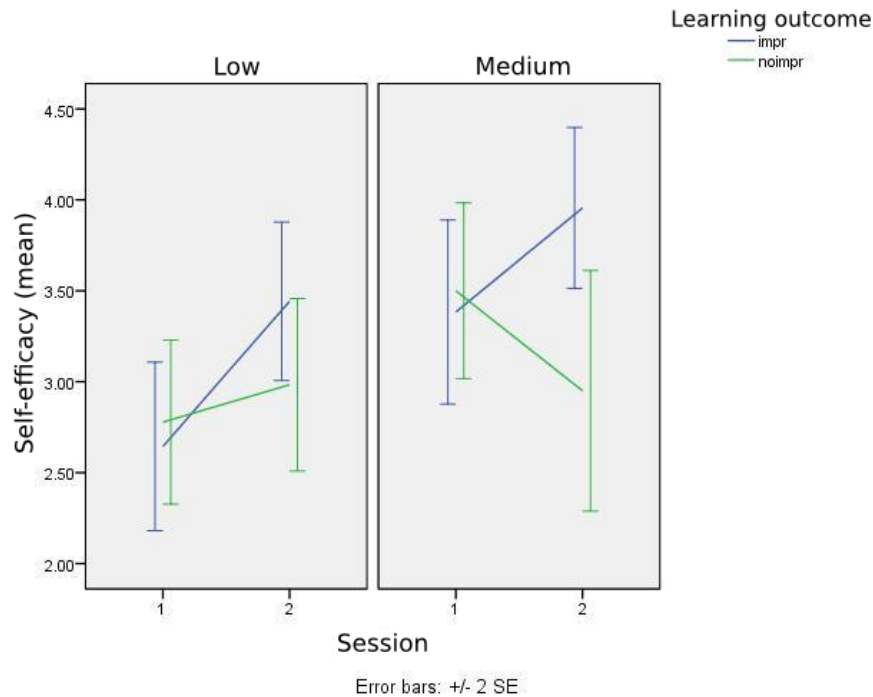


Figure 4.9: Self-efficacy as related to initial proficiency and learning gains. Left: low proficiency learners report significantly higher self-efficacy in the last session, with no learning outcome effect. Right: for medium proficiency learners, those who improve report significantly higher self-efficacy in the last session, with a significant difference in the last session between learners who improve and those who do not improve.

proficiency (see Figure 4.11), $F(1,48) = 9.45$, $p = .003$, $\eta_p^2 = .165$, showing that the medium proficiency group found practice to be significantly easier ($M = 4.12$, $SD = 0.67$) than the low proficiency group ($M = 3.5$, $SD = 0.74$), which is a reasonable finding.

The analysis also finds a significant interaction effect for proficiency and practice condition, $F(1,48) = 4.9$, $p = .032$, $\eta_p^2 = .093$. Post-hoc analysis reveals, first, that within the low proficiency group, learners in the NOCF condition report more difficulty ($M = 3.14$, $SD = 0.46$) than those in the CF condition ($M = 3.81$, $SD = 0.81$). Second, for learners in the NOCF condition, low proficiency learners report higher difficulty levels than medium proficiency learners ($M = 4.40$, $SD = 0.82$).

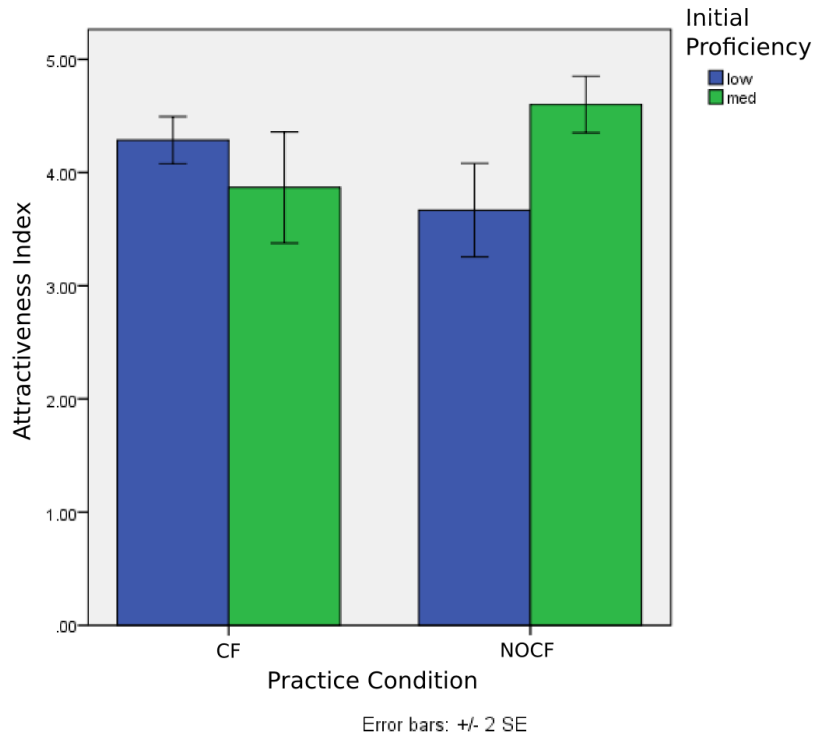


Figure 4.10: The influence of factors practice condition and proficiency on the variable attractiveness.

4.7 Discussion and conclusions

In Chapter 3, we investigated learners' affective states and user-system interactions on a micro-scale and suggested that CF can be provided in a CALL-ASR environment without negatively impacting learner affect. The present study built on the previous one by investigating learner affect in the same self-study environment with a larger number of participants and a more comprehensive evaluation that added an analysis of learning gains, a post-practice evaluation and their relationships with in-practice affect.

In answer to our first research question, it seems that practice with the system generally leads to learning gains and positive affective experiences. Part one of the present study indicates that learners enjoyed and were willing to practice with the system in both conditions with CF and without CF, reaching the same conclusion as our previous work examining changes in affect within practice sessions. This finding is corroborated by the results of the

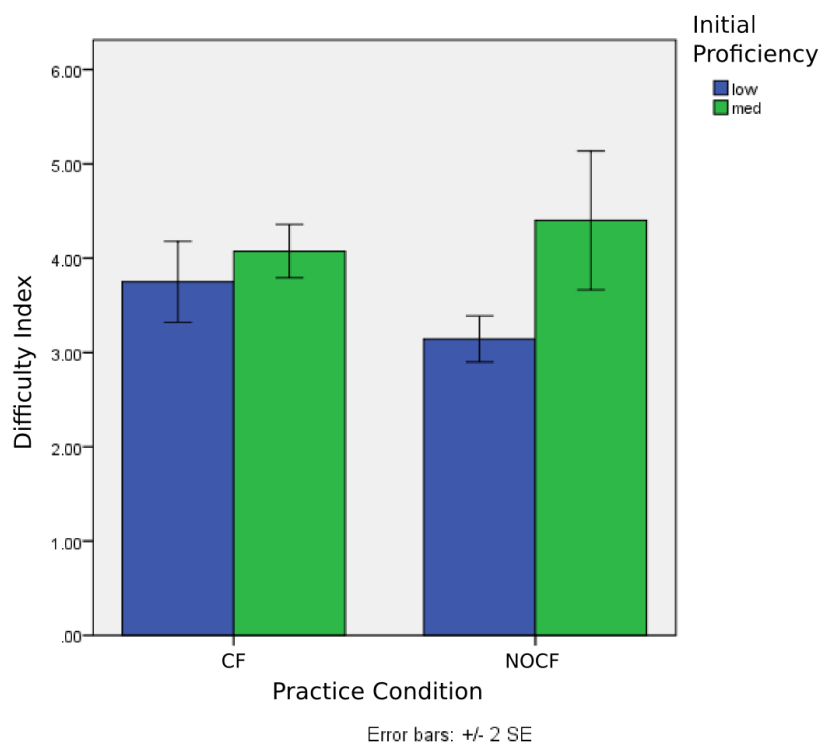


Figure 4.11: The influence of initial proficiency and CF condition on the variable difficulty.

post-practice questionnaire: when learners reflect on and evaluate their practice experience, they express positive evaluations of the activity. This applied to both conditions investigated, CF and NOCF. Here, it seems worth noting that we also performed a separate analysis including only the data of the group who completed five practice sessions, which produced similar findings and suggests that learners experience the system positively even after a longer period of time. Equally important are the results of the proficiency tests. The analysis found a significant effect for the factor session, indicating that learners' oral productions improved in word order accuracy over the course of the experiment, but immediate feedback as provided in the CF condition did not lead to a significant advantage. To answer our second research question, there appears to be no significant impact of automatic CF on learning gains and in-practice affect in the present learning context.

An important aim of this study was to assess students' reactions to the automatic CF provided by the ASR in relation to their learning experience.

Feedback is often requested by learners yet providing CF on oral productions effectively in the classroom is challenging and may cause anxiety. Our previous investigation suggested that this may not apply in ASR-based self-study CALL and the results of the present study replicate this finding, showing that CF can be provided in the current oral practice context without the negative impact sometimes claimed or reported in the literature (e.g. Agudo, 2013; Loewen et al., 2009; Truscott, 1999). Thus, our findings regarding CF and affect seem to be in line with MacIntyre, Burns and Jessome's observation (2011) that "the methods of delivering error correction, whether it is welcomed by the student in context and the relationship with the teacher converge to affect WTC" (p. 89).

In part two, we shifted the focus from the learning environment to individual learner variables. An analysis of the variation in initial V2 proficiency levels and learning outcomes revealed that meaningful sub-groups of learners could be created with large enough group populations to support an analysis of the relationship between these variables and learner affect. The results show that despite having comparable overall proficiency levels (CEF A1 and A2), learners had initial V2 accuracy scores that spanned a wide range. The analysis also reveals that there was variation in the learning gains: most learners improve, but a near equal number of learners (26/56) do not, which is in line with the finding that V2 is difficult to master (Jordens, 1988), as it does not alter the meaning of a sentence and therefore is less easily noticed by learners.

The difference in outcomes observed provided an opportunity to look at how in-practice affective states are related to learning gains. Interestingly, separating learners based on their initial proficiency reveals that the relationship between affect and learning gains is different for different proficiency levels. Learners with low initial V2 accuracy experienced practice with the system positively in the first and last sessions, regardless of whether they actually improved on V2 accuracy or not. On the other hand, learners with medium initial V2 accuracy who did not improve reported lower levels of enjoyment, willingness to practice and self-efficacy in the last session, while medium proficiency learners who improved reported positive affect in both sessions. A possible explanation for the affective differences observed between the low and medium proficiency groups is that the medium proficiency learners had more resources to gauge their progress and became more critical if they felt they were not improving. The self-efficacy results showing lower values for the medium proficiency learners who did not improve might be interpreted along these lines.

Differences between the low and medium proficiency groups were observed also in the way they valued CF. Although CF did not impact in-practice affective states, after practice, low proficient learners in the CF condition found practice more attractive and easier. To answer our third research question, initial proficiency and learning gains are indeed related to affective experiences. These findings seem to reinforce the idea that CF is particularly valuable for low proficiency learners, as observed in (Penning de Vries et al., 2014a). In that study, CF was positively evaluated by L2 learners, but it also led to

a significant improvement in oral skills for low-proficient learners, while the latter did not emerge from the present study. This adds to the body of research indicating the complexity of the relationship between CF, L2 proficiency and learner characteristics (Penning de Vries, 2015).

Together, the proficiency tests and positive experiences reported in the in-practice and post-practice questionnaire data indicate that a CALL system with ASR and automatic CF can provide conditions supporting oral L2 practice in a comfortable environment. CF may not always be necessary to improve oral proficiency, but it is valued by L2 learners, especially those with lower proficiency, and it does not always have a negative impact on affective states. These results should be encouraging to developers of CALL-ASR systems. It could also be interesting news for teachers, as grammar practice is typically not perceived as enjoyable (Jean & Simard, 2011), yet in this context this turned out to be the case.

CHAPTER 5

Investigating motivational diversity through a possible
selves approach: a computer application for adult L2
learners of Dutch

This chapter is to be submitted as

Bodnar, S., Cucchiaroni, C., Strik, H., & van Hout, R.W.N.M. Investigating motivational diversity through a possible selves approach: a computer application for adult L2 learners of Dutch (to *Computers in Human Behavior*).

Acknowledgements

This work is part of the research program Feedback and the Acquisition of Syntax in Oral Proficiency (FASOP) NWO-Dossier Number 360-75-010, which is funded by the Netherlands Organisation for Scientific Research (NWO).

Abstract

Language learners are known to have different motivations for learning a second language (L2). In the classroom, teachers can draw on these different motivations to personalise practice and create more engaging learning environments. One problem, however, is that surveying motivational profiles and developing tailored materials may not be feasible with teachers' limited time. Computer programs that gather motivational profiles may be able to provide tailored language practice without extra burden for the teacher. An important first step in this line of research is to explore the motivational diversity of target groups of learners to gain insight into whether such a computer program is necessary or feasible. The present study surveyed the motivational characteristics of a group of adult Dutch language learners ($N = 59$) using a possible selves approach. A computer activity was developed in which learners browsed through a set of cards describing potential L2-related futures, selected cards that were significant to them, sorted the cards according to their desirability and rated the top-desired cards on self-efficacy and plausibility. The desirability scores generated from this activity were submitted to an inverted factor analysis to explore participants' motivational heterogeneity and profiles. Sub-groups based on desired aspects of learners' ideal L2 selves were identified and similarities and differences between representative motivational profiles were explored. In a second step, subgroups of learners having similar ideal L2 self characteristics were examined for heterogeneity of self-efficacy and plausibility beliefs. Results show 1) that learners overlap in their motivations for learning Dutch, but also suggest that motivation for learning a second language can be diverse, and 2) that self-efficacy and plausibility beliefs concerning the most-desired aspects of learners' ideal L2 selves were homogeneous. The implications of this finding for developing personalised language instruction in the context of computer-assisted language learning environments are discussed.

Keywords: L2 motivation, possible selves, instrument, method

5.1 Introduction

It is generally accepted that learning situations are most effective when they are personalised to the characteristics of the individuals they teach. Williams and Burden (1997) express this view well when they state “every learning experience should be seen within the context of helping learners to develop a sense of personal identity and relating that to realistic future goals, i.e. learning should be personalised as far as possible” (p. 36). However, there are practical obstacles to realising widespread implementation of personalised learning: teachers are busy individuals, often under high stress, who have limited time for lesson preparation (Carless, 2003; Crookes & Arakaki, 1999). Taking the time to become familiar with individual learners’ characteristics and developing tailored materials may not be feasible for many learning situations.

This issue presents researchers in language learning and technology with an interesting research direction: how can technology be used to deliver personalised learning experiences? One approach has been to combine autonomous learning approaches with resources on the internet in self-study language labs (e.g. Raby, 2007). A second approach has been to work closely with learners during the software design and development process, by eliciting their emotional responses to the software and suggestions for improvement (Colpaert, 2010). A third option that has sometimes been investigated in the CALL literature is the development of computer-based activities that stimulate learners to reflect on their study habits and processes (M. Chang, 2005). A fourth possibility is the tutorial Computer Assisted Language Learning (CALL) approach (Heift & Schulze, 2015). Found mostly in the research lab, but with some implementations in authentic learning environments (e.g. Heift, 2016), tutorial CALL uses a combination of artificial intelligence technologies to provide an individualised tutor-like experience to language learners with features such as adaptive feedback.

What has not been investigated much, but what in our view constitutes an innovative research direction, would be the combination of the latter two approaches, self-reflective activities and tutorial CALL, to address affective individual differences. Tutorial CALL systems typically employ technologies to adapt aspects of practice related to learners’ cognitive characteristics, but some exceptions which incorporate affective individual differences exist. For example, Heilman, Collins-Thompson, Eskenazi, Juffs, and Wilson (2010) developed a vocabulary tutoring system that used these technologies to personalise reading exercises by selecting passages that matched learners’ interest profiles, making this one of the few CALL systems that adapt to learners’ affective characteristics.

One affective dimension of learning that is widely accepted as both varying with individuals and having an important influence on learning is motivation. The motivation to learn a second language has been approached from a variety of perspectives, including *integrative orientation*, reflecting learners attitudes towards the target language speakers and culture (Gardner, 1985b), *instrumen-*

*tal*ity, reflecting expected pragmatic benefits, (e.g. getting a job) (Gardner, 1985b), *intrinsic motivation*, describing the degree to which an activity is inherently enjoyable or interesting, and its counterpart *extrinsic motivation* (Noels, 2009), representing the extent to which an activity is interesting because it serves some other, external purpose, and present self-concept and *ideal* and *ought-to future self-guides*, which capture how the learner sees themselves at present and their thoughts concerning the kind of person they could become in the future (Dörnyei, 2005, 2009a).

An interesting possibility, then, could be the integration of computer-based reflective activities that elicit motivational information with tutorial CALL applications that use this information to tailor the learning system, resulting in personalised practice without extra burden for the teacher. However, this idea raises a number of important questions about eliciting and employing motivational information from learners, including concerns about data (what kind of information is useful?), method (how should we elicit this data?), feasibility (what methods could we use to take this data and adapt computerised practice), and effectiveness (what is the effect of personalising language practice based on motivational characteristics, in terms of affective experience and learning outcomes?).

An important first step towards building a tutorial CALL system that personalises instruction is to investigate whether personalisation is necessary and how difficult of a problem it is. In a relevant and relatively new direction in L2 motivation research, some researchers have begun investigating motivational types (e.g. Csizér & Dörnyei, 2005b; Papi & Teimouri, 2014). These studies aim to “explain how such motivational variables fit in the students’ motivational configurations and how these configurations result in different emotional, motivational, and linguistic states” (Papi & Teimouri, 2014, p. 519). Advances in this direction could reveal how different motivational strategies are more (or less) effective for certain motivational profiles. Although at present relatively few studies have been completed, continued research into learners’ motivational types and their diversity among learners could provide valuable insights. Such research could provide information on the contents of learners’ motivational profiles, the commonalities and differences between the profiles, and the proportional abundance of the different profile types. This information could in turn provide suggestions on how one could approach creating a tutorial CALL environment that adapts to learners’ motivational characteristics.

The present study is a first step towards addressing these questions and challenges in a CALL context. In section 2 of the present study, we review the language learning motivation literature, as well as the educational psychology and motivational psychology literature, to inform questions regarding motivational information and the techniques that might be used to elicit it. In section 3, we then explain how this information has been used to design and develop a computer activity for obtaining information regarding learners’ desired future selves, which could serve as a resource for personalising instruction. In section 4, we report on a preliminary use case of the application in which learners of

Dutch completed the activity at home, prior to visiting a university language laboratory to participate in an experiment. Drawing on the data from the learners, we investigate the contents of language learners' ideal L2 selves as well as the prevalence of different types of ideal selves, to obtain suggestions for implementing personalised practice. Section 6 presents the results, showing the population to consist of three major types, but to also be diverse. We relate this information to previous work on motivational types, consider how this information could be used in motivational strategies, and reflect on the implications of our findings for the feasibility of developing personalised content. Section 7 offers a general conclusion and future directions.

5.2 Background

This section provides an overview of the theoretical framework used in the present study. Following this, we review the language learning literature on studies of motivational types before presenting our research focus.

5.2.1 Theoretical framework

The theoretical framework adopted in the present study is the L2 Motivational Self System (Dörnyei, 2005, 2009a) (L2MSS). This framework draws on theoretical concepts related to the self from motivational and educational psychology. Therefore, before describing the L2MSS, we provide a brief overview of the concept of 'possible selves' and 'L2 self-guides', on which the L2MSS is based.

The term *possible selves* (Markus & Nurius, 1986) is used in the motivational and educational psychology literature to refer to cognitive representations people have about the kind of person they might become in the future. To give an example, if a language learner thinks or imagines themselves as *a highly fluent Dutch speaker*, we can say that this image of themselves constitutes one of their possible selves¹. While an individual could conceivably imagine themselves in the future in any number of ways, according to Self-Discrepancy Theory (Higgins, 1987), certain possible selves come to serve as important reference points for the individual which guide their behaviour, referred to as *self-guides*. Higgins (1987) proposed that perceived discrepancies between a person's present self-concept and an imagined future self can induce affective states in an individual. In particular, Higgins, Klein, and Strauman (1985) provided empirical evidence that showed how perceived discrepancies between the current self-concept and an *ideal self* could result in dejected states such as sadness or depression, while discrepancies between the current self-concept and an *ought self* could result in agitated negative states such as anxiety. Applied to language learning, these theories suggest that, for example, the experience of imagining oneself failing to watch movies in an L2 one is interested in without

¹We will use the terms possible selves, possible future selves and future selves interchangeably for stylistic reasons.

subtitles might evoke disappointment, while picturing oneself struggling to find the right words during a conversation in an L2 may evoke anxiety or a desire to avoid encountering that future.

With the L2MSS, Dörnyei (2005, 2009a) has drawn on the concepts of possible selves and self-discrepancy theory to propose a self-based theoretical account of motivation for the domain of language learning. The system is composed of two self-guide components and a third concerning the learners' affective experiences in their immediate learning environment. The first self-guide component is the *ideal L2 self*, which is used to refer to a learner's vision of the kind of L2-speaking person they aspire to become. The second self-guide component is referred to as the *ought-to self*, which is associated with undesirable L2-related visions of oneself that one hopes to avoid. With respect to motivation, the ideal L2 self is associated with a promotion (or approach) focus, while the ought-to L2 self is associated with a prevention focus (Dörnyei, 2009a). To continue the previous example, learners who aspire to watch Dutch films without subtitles in the future (an aspect of their ideal L2 self) might feel motivated to study Dutch to approach their ideal self. On the other hand, learners who imagine themselves losing face if called upon in class by the teacher might feel motivated to do their homework (an aspect of their ought-to L2 self) to prevent that future.

A number of medium and large-scale studies have shown that ideal L2 selves are positively linked to the motivation to learn an L2 (e.g. Al-Shehri, 2009; Csizér & Lukács, 2010; Islam, Lamb, & Chambers, 2013; Kim, 2012; Taguchi, Magid, & Papi, 2009), while relatively few studies have investigated the conditions under which a possible self can become a self-guide with motivational impact. Hessel (2015) is one of the most recent studies to aim at a more detailed understanding of how the attributes of a possible self can influence its motivating effects (but see also MacIntyre, Mackinnon, & Clement, 2009). Participants were German university students with an upper-intermediate to advanced knowledge of English. Noting shortcomings with previous instruments, Hessel (2015) developed a new instrument that asks learners to provide information on a self's desiredness, accessibility, plausibility, discrepancy with present-self, and as a criterion measure, the effort exerted towards achieving that self. Analysis proceeded on an item basis, using multiple regression to examine the relative contribution of each of the possible self attributes in predicting the effort exerted in pursuing that possible self. Results from the analysis reveal that the accessibility property of a possible self, i.e. how often a possible self is imagined, was the largest predictor of effort. The self's discrepancy with the current self-concept, which has been theorised to be the main predictor of self system's motivation, was found to be the second largest predictor, followed by the desiredness of the self. As Hessel (2015) concludes, the finding that accessibility is "by far the strongest and most direct predictor . . . of self-motivated engagement in L2 learning" is surprising, given the theoretical importance placed on self-discrepancy. At present, it seems that accessibility, self-discrepancy and desirability are attributes of possible selves which most

predict their motivational capacity, but that more empirical research is still needed into the conditions under which possible selves become motivating (Taylor, 2013).

The third component in the L2MSS system is the *L2 learning experience* which is concerned with peoples' affective experiences in their immediate language learning environments. Dörnyei (2009a) lists “the impact of the teacher, the curriculum, the peer group, the experience of success” (p. 46) as factors influencing the affective experience. As an example, in their questionnaire, Csizér and Kormos (2009) operationalised the L2 learning experience as “the extent to which students like learning English” (p. 111).

At present, the L2 learning experience has not been theorised to the same extent as the self-guide components of the L2MSS (Hessel, 2015; Taylor, 2013). The lack of an extensively theorised interface between the self-guides and the L2 learning environment could make an L2MSS-based approach to personalising CALL practice seem questionable. However, we believe there are good reasons to assume that the relationship between self-guides and the L2 learning experience is a strong one. A detailed examination of theory relating to this interface is beyond the scope of the current paper, but we briefly sketch one approach for how this might work. According to Carver (2006),

Goal pursuit means having a goal, assessing where one is with respect to it, and taking steps to reduce the sensed discrepancy. This idea can be applied to motor-control goals such as reaching and grasping; it can also be applied to goals that are more abstract (e.g., being honest, being productive), and goals that are continually moving and evolving targets (e.g., developing a research career, having a sound marriage, raising children with the right kind of personal values). (p. 106)

Following this definition, L2 self-guides can be seen as a kind of abstract goal that is continually changing and evolving. Indeed, possible selves (Markus & Nurius, 1986) and the ought self-guide (Higgins, 1987) are mentioned as examples of goals in his discussion of self-discrepancy and the resulting impact on affective state. Carver's model (2003; 2006) proposes that individuals have a feedback system that monitors the discrepancy between a person's current state and future goals or threats, and which motivates their behaviour to approach or avoid these future states. A second function of the feedback system (a second “layer” in Carver's model) is to sense the rate of progress towards a goal. As Carver (2006) explains,

The sensed rate of progress of the first system is compared to a criterion, and the result (the “error” between the two) is experienced as affect. If the rate is below the criterion, negative affect arises. If the rate is at the criterion, the person is affect-free. If the rate exceeds the criterion, positive affect arises. In essence, the idea is that positive feelings mean you're doing better at something

than you need to (or expect to), and negative feelings mean you're doing worse than you need to.

According to this theory, when individuals perceive that they have made progress towards a desired future goal such as their ideal L2 self, the reduction in discrepancy between the current state and future goal is experienced as positive affect. When individuals perceive that they have distanced themselves from an undesired future state such as their ought-to L2 self, the increase in discrepancy is experienced as a reduction in negative affect.

These mechanisms, based on perceived rates of progress to and away from important imagined future states, provide a basis for why we might expect learning environments that are aligned with learners' ideal and ought-to L2 selves to influence affective learning experiences, and in turn impact learning.

5.2.2 Research into self-concept and motivational types in language learning

Studies which investigate motivational types shift the focus from variables to persons. Only a few studies attempt to profile persons using characteristics of their L2-related self-concept (Csizér & Dörnyei, 2005b; Irie & Ryan, 2015; Papi & Teimouri, 2014). The first two studies are relevant, as they investigate future-oriented motivational types. The third study is relevant for its look at the types of present L2 self-concepts and for its methodology.

The study by Csizér and Dörnyei (2005b) had the aim of identifying motivational types in a large-scale survey of Hungarian public school students ($N = 8,593$, aged 13 - 14) using cluster analysis. The survey completed by the students recorded their attitudes towards five different languages and target cultures (e.g. English / US, English / UK, German, French, etc.) and obtained ratings along the motivational dimensions of integrativeness, instrumentality, vitality of the community, attitudes towards L2 speakers and cultural interest. These results were then submitted to a cluster analysis. In a second step, a statistical analysis was run comparing each motivational types' intended effort to learn the L2 and their preferred language choice. Results from the study suggested the presence of four distinct motivational profiles, which were interpreted in terms of the L2MSS. The profiles are characterised as having two complementary aspects: 1) an image in which we appear "personally agreeable (associated with positive attitudes toward the L2 community and culture)", and 2) an image in which we are "professionally successfully (associated with instrumental motives)" (p. 637). Concerning the groups that emerge, Group 1 is reported to have both a lack of appreciation for the L2 community and culture and for the potential professional benefits of knowing an L2. Group 2 is found to have more positive attitudes towards L2 speakers, cultural interest and vitality, but view English as irrelevant to their future professional life. Group 3 is the opposite: they lack an appreciation of the L2 culture and community, but have stronger positive attitudes towards the potential usefulness of English in their

career. Group 4 is positive on all of the motivational variables, including both cultural and professional aspects of their ideal L2 self. On the criterion variable of intended effort, results indicate that Group 1 was lowest, with Groups 2 and 3 following in that order, with Group 4 reporting highest intended effort. The differences between Groups 2 and 3, which are lacking complementary aspects in their motivational profiles, are explained by Group 3's strong appreciation for the professional aspect of their ideal self.

Papi and Teimouri (2014) investigated the motivational types of middle and high-school English learners in Iran ($N = 1,278$, ages 13 - 18), also using cluster analysis. In this study, a questionnaire was used to elicit key components of the L2MSS, including global measures of the ideal and ought-to L2 self-guides, as well as constructs typically associated with those guides, such as attitudes towards the L2 community and family influence. They obtained data on three self-report criterion variables: perceived language proficiency, motivated learning behaviour and language anxiety. Results from the questionnaire were first submitted to a cluster analysis. In a second step, a statistical analysis was used to compare the groups' motivational profiles, as well as their ratings on the criterion variables. In a third step, the study categorised learners into promotion-oriented students and prevention-oriented students and investigated differences in both the constructs making up the motivational profiles and the criterion variables. Application of the cluster analysis produced five motivational types. Group 1 was characterised as not seeing any relevance of English in their future life. Both their ideal L2 self and ought-to L2 self, together with their associated constructs, were found to be the lowest of the groups; their moderate anxiety suggested they had concerns about passing a mandatory English course. Group 2 appeared to find learning English enjoyable but not particularly relevant. They had a weak ideal self and weak ought-to self, indicating that they had difficulty seeing the relevance of English in their future. They possessed moderate attitudes towards the L2 culture community, which were interpreted to likely be a product of their enjoyment of their attendance at after-school private language institutes. Group 3 recognised English as an important part of their work-related future, but appeared uninterested in the L2 culture or community. They had a moderate ideal L2 self, weak attitudes towards the L2 culture or community, yet a strong instrumental-promotion orientation. They had a moderate ought-to self and a strong instrumental prevention orientation. Group 4 possessed strong ideal L2 selves and had positive attitudes towards L2 culture, the native speaker community and instrumentality promotion, with the researchers interpreting their motivation as well-rounded. Their ought-to self, however, was low, as were the associated constructs. This was interpreted to mean that they were unconcerned about potential negative outcomes. Finally, Group 5 had a strong ideal self and associated constructs, appearing to have well-rounded motivational characteristics including social and professional aspects. They also had strong ought-to selves, indicating they had concerns about avoiding potential undesirable outcomes. The group was found to be similar to the

highest motivated group in the study by Csizér and Dörnyei (2005b).

The two studies we have presented so far have similar methodological approaches. Although questionnaires are one of the most frequently used methods for eliciting possible selves data, as Packard and Conway (2006) show in their review of methods in the broader psychology literature, they are not the only method available. In broader psychology and in language learning, researchers interested in eliciting data related to possible selves have also frequently employed interviews, and sometimes used less-frequent methods such as narrative (e.g. Thompson & Vásquez, 2015), visual methods, dramas, and even video games (King, 2011). Below, we look at one study which uses a less frequently used method, card sorting, which is also the technique used in the present study.

Irie and Ryan (2015) employed a dynamic systems perspective to investigate how learners' L2-related present self-concept changed in relation to a study abroad experience. The study was conducted with 20 Japanese students of English who went overseas for a five-month study-abroad experience. The study is noteworthy for substituting the questionnaire approach with less-frequently used Q methodology (S. R. Brown, 1980, 1993; Watts & Stenner, 2012). Q methodology (or simply 'Q') employs a card sorting activity to elicit participant data. Learners were asked to review and sort a collection of index cards on which short texts describing self-concepts were written². The criterion for the sort was the perceived self-similarity of the statements (i.e. "How descriptive is this statement about your view of L2 learning and L2 use?", Irie & Ryan, 2015, p. 349). Texts on the cards were drawn from existing questionnaires related to the L2MSS (e.g. MacIntyre, Mackinnon, & Clement, 2009; Taguchi et al., 2009) so as to maintain links with established motivation research. Examples of the items included desirable statements, such as "Interacting with foreign people in English is fun for me" (Irie & Ryan, 2015, p. 355), as well as undesirable statements such as "I don't have the right personality for learning English" (ibid., p. 355). Learners completed the card sort once before and again after the study abroad experience. Using an inverted factor analysis technique, Irie and Ryan (2015) compared the types of self-concept present at time 1 and 2. The results from the study are interesting because they point to change in self-concept and identify different trajectories. At time one, prior to travelling abroad, it was found that the self-concept of the participants could be best described with one type, that of a "naive optimist" (p. 354) who is confident, highly-motivated, but unrealistic. When data was collected at time 2, after five months of being overseas, three self-concept types emerged. First, some students seem to have had decreases in efficacy and plausibility. These students were described as "shell-shocked doubter[s]" (p. 354). Second, other students seem to have experienced similar changes in efficacy and plausibility, but possessed an ought-to self that tempered their despair and thoughts of giving up. Irie and Ryan described this type with the label "duty-bound learner" (p. 356). Third, the study observed some individuals who returned as "comfortable user[s]"

²At time 2, a computerised version was also used

(p. 355), with their self-concept similar to how it was at time 1, suggesting these individuals managed to make friends, feel comfortable, enjoy English, and remain positive.

5.2.3 Summary and the present study

According to the L2MSS (Dörnyei, 2005, 2009a) learners can have ideal and ought-to L2 selves which are mental representations of the kind of L2-speaking person they hope to become and to avoid becoming in the future. Generally speaking, learners feel motivated to approach their ideal L2 selves, and motivated to avoid their ought-to selves. However, the motivation can vary, according to different attributes of the self-guides, such as their desiredness, discrepancy with the present self, accessibility and plausibility (Hessel, 2015). These factors interact with each other, together with learners' affective experiences in their learning environment, to influence the effort learners expend and learning outcomes. Although the interface between the ideal and ought-to L2 selves and the L2 learning experience has not yet been comprehensively theorised in the L2 motivation literature, a self-regulation mechanism proposed by Carver (2006) seems to integrate well.

As a first step towards investigating an possible selves approach to personalisation in a tutorial CALL system, we were interested in developing a computer activity for eliciting self-guide data. First, we were interested in developing a tool that could tell us about the diversity of learners' self-guides, as we thought this could inform us as to whether personalisation was necessary: if learners' self-guides are homogeneous, than personalisation technology may not be needed; if they are diverse, than personalisation technology could benefit learners and it would be worthwhile developing adaptive technology. Second, we were interested in a method that could obtain information about the contents of learners' self-guides to reveal motivational themes that may exist and potentially suggest directions for personalising practice. Our research questions were:

1. What might a computerised tool that elicited information on learners' L2 self-guides look like?
2. What motivational types would emerge from application of this tool to language learners?
3. How could integrating such a tool into a CALL system help to enable personalised practice?

In the next section, we describe how we developed a computer application for eliciting L2 self-guide data with the aim of facilitating the personalisation of instruction.

5.3 A computer activity to elicit information concerning L2-related self-guides

In designing a tool to elicit L2-related self-guide data, we had a number of criteria in mind. First, in order to obtain directions from the learner for personalising practice, the tool should aim to obtain data on the *contents* of a learner's self-guides. Second, the tool should be suitable for automating in a computerised environment, to facilitate personalisation while leveraging technology to minimize demands on educators. Third, the method should be engaging to learners and generally be a positive experience. Finally, the method would ideally elicit information about different properties of a self-guide (e.g., desirability, plausibility, discrepancy with presents self-concept, etc.) to enable more research on the motivational conditions under which a possible self becomes motivating (Hessel, 2015; Taylor, 2013).

In their review of methods for eliciting possible selves data, Packard and Conway (2006) articulate a number of dimensions, at least two of which are relevant for designing a computerised self-study activity. First, whether a method is structured or unstructured affects its potential for automatisation as a computer program. The structured methods, namely surveys, structured interviews and card sorting techniques are more easily automated than the less structured methods of narrative, visual and drama. Second, whether a method is closed or open also affects its potential to be automated. Closed methods refer to an approach where learners interact with a fixed set of pre-existing items, typical of Likert-scale questionnaires. In open methods, participants respond to a question or instructions by generating their own content, most often with text but, visual or game mediums are also possible. Because of the challenges of natural language processing, closed methods are most suitable for automating.

In the present study, our approach was to develop a guided card sorting activity. In card sorting methods, typically short texts describing a variety of future selves are written on index cards; participants use the cards in a variety of selection and sorting activities (Kerpelman, 2006). Although commonly used in the broader psychology literature (Packard & Conway, 2006), card sorting is only starting to attract attention in the language learning literature (e.g. Irie, 2014; Irie & Ryan, 2015; Lamb, 2013; Pemberton & Cook, 2012). Compared to structured surveys, card sorting is effective for eliciting configurations of motivational variables, showing how a learner considers variables in relation to each other (Packard & Conway, 2006) and is said to be a good reflective activity that encourages more engagement from learners (Irie, 2014; Pemberton & Cook, 2012). Although some commercial card-sorting platforms exist (*CardSorting.net*, 2016; *OptimalSort*, 2016), we decided to develop our own activity because we were interested in creating a dedicated reflective activity for learners as a future component of a tutorial CALL system. Our aim was to provide a guided, tutorial-like interaction and pace, and to improve on user interfaces of current computer-based card sorting

applications by providing an experience that is similar to working with paper index cards. Below, we explain the design, present the activity and describe the data that is generated.

5.3.1 Overall design of the activity

The application aims to provide learners with a comfortable and engaging environment that elicits data on different possible futures. The activity is implemented as an interactive web page and includes instructions and trial activities designed to familiarise the learner with how they can interact with the system. It is modelled after Shepard and Marshall's Possible Selves Mapping Interview (Shepard & Marshall, 1999), and centres on card selecting and sorting activities.

5.3.2 Selection of future self cards

As mentioned, the goal of the application was to prompt learners to reflect and elicit information related to self-guides that could be used as a system resource for adapting practice. The focus is on probing the nature of learners' future perspective on Dutch as an L2. Therefore, we developed items that they were descriptions of hypothetical selves a learner might become in the future. The cards used in the activity contain short statements describing different characteristics of one's self related to language learning (e.g. "I can use Dutch to get a better job", "I know Dutch slang and can use it well"). The texts describing the selves were worded in the present tense because of the design of the exercise: learners were asked to imagine that the self on the card is true, with the intention of stimulating an actual-future comparison.

Ideal and ought-to L2 self-guides are both considered to have important roles in influencing motivation, choices made, and expended effort in the learning situations. Therefore, in developing the cards for the activity, we decided, initially, to create cards to investigate both potential desired and undesired aspects of learners' futures (although see below). Similar to Irie and Ryan (2015), the majority of texts for the cards are based on items from previously published questionnaire instruments (e.g. MacIntyre, Mackinnon, and Clement 2009) and adapted for the present activity. We worked iteratively, by initially creating a large number of cards from a pool of questionnaire items. Through discussions between the first and second authors, language learners, and educators, we refined this pool to a set of 48 cards. The items describing desirable and undesirable possible futures adapted from questionnaires, comprising 41 of the 48 items, corresponded to the following categories:

- *Attitudes towards L2 community*, concerning futures where learners interact and get along well with Dutch people (4 items: 3 desired, 1 undesired).

- *Cultural interest*, describing futures where learners have access to Dutch media (4 items: 2 desired, 2 undesired).
- *Integrative orientation*, concerning different futures in which the learner interacts with or becomes similar to Dutch people (4 items: 4 desired).
- *Instrumental orientation*, describing potential pragmatic benefits of learning Dutch (4 items: 2 desired, 2 undesired).
- *Interest in foreign languages*, describing futures related to use and knowledge of foreign languages in general (4 items: 2 desired, 2 undesired).
- *International posture*, depicting futures where the learner has opportunities to learn about foreign cultures and have contact with people from around the world (2 items: 2 desired).
- *Linguistic competence*, concerning futures in which the learner is an all-around competent Dutch user (4 items: 2 desired, 2 undesired).
- *Linguistic proficiency*, describing potential futures in which the learner has become proficient in Dutch (4 items: 2 desired, 2 undesired).
- *Milieu*, describing futures in which the learner has pleased others by learning Dutch or is appreciated for doing so (4 items: 2 desired, 2 undesired).
- *Willingness to communicate*, concerning futures in which the learner feels confident and comfortable speaking Dutch (4 items: 2 desired, 2 undesired).
- *Fear of assimilation*, describing fears regarding the possibility that learning and using foreign languages might lead to the loss of the native language and culture (3 items: 3 undesired).

Based on informal discussions with language learners and research investigating L2 learners expression of emotion and slang (e.g. Dewaele, 2010), an additional seven new items were added to capture potentially desirable and undesirable futures that were not represented in the reviewed questionnaires:

- *Self expression*, concerning future states in which the learner has the capacity to express themselves (3 items: 3 desired).
- *Self-efficacy*, describing a future in which the learner is capable of communicating without assistance (2 items: 1 desired, 1 undesired).
- *Peer contact*, describing the possibility of spending time with other Dutch learners in the future (2 items: 1 desired, 1 undesired).

5.3.3 Flow of the activity

The activity consists of a series of hyperlinked interactive web pages. There are three different types of pages: non-interactive (standard) instruction pages; interactive instruction pages; interactive activity pages. A progress bar is provided at the top of the screen to provide feedback.

Login and instructions

After the learner logs in to the system, they are brought to a home page where they can launch the activity. The initial pages for the activity aim to set the scene and encourage learners to reflect on their futures.

Small steps to familiarise the learner with the user interface are taken, using a combination of standard and interactive instruction pages. First, an interactive instruction page presents learners with example cards and allows them to browse the cards. The aim here is to become familiar with reading the cards and the type of texts written on the cards. Instructions on this page are as follows:

Imagining yourself in the future

In the list below there are 8 cards. Each card describes something that might happen to you in the future. Most cards are related to learning Dutch, but there are other ones too (for example, about the country you live in). Try reading the cards now. Use the arrows to look through the cards.

The second page is an interactive instruction page which is intended to allow the learner to practice moving the cards while making judgments about their desirability. This page makes use of three cards, specially chosen so that they would provide good examples of wanted, unwanted and irrelevant futures and are not part of the card deck used in the activity. For example, the first section introduces learners to desired cards (see Figure 5.1). On the left, a card with the text “I am in good health” is shown, and on the right a “card drop area” is presented. The learner is asked to review the card and to drag it to the “Likes” card drop area.

Card selection

In the next step, a card selection activity is presented. Learners are asked to browse through the deck of 48 cards and select cards with future selves they either desire or wish to avoid. The order of the cards is randomised. Learners select a card by dragging it to a card drop area. Two card drop areas are provided, for desired and undesired futures. For cards that are irrelevant, the system instructs learners to simply skip them, by leaving them where they appear in the card browsing area.

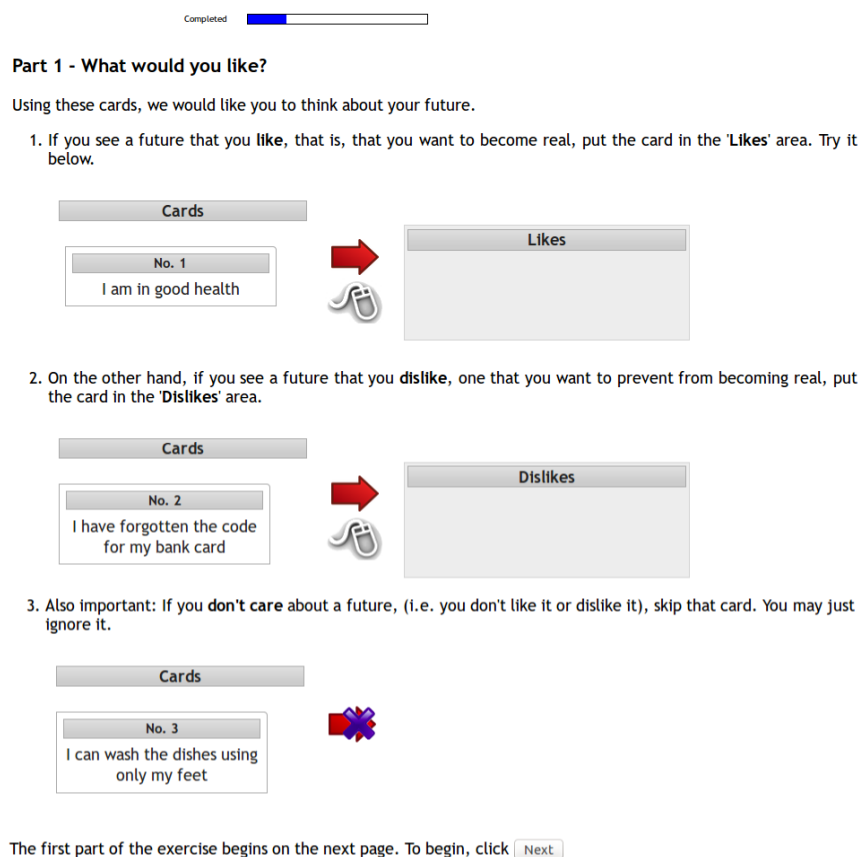


Figure 5.1: A screenshot of an instruction page. Instruction pages help to familiarise learners with the task before completing the activity.

Sorting by desiredness

The next step consists of a card sorting activity where learners are asked to consider how attractive they find the different futures they have chosen (see Figure 5.2). Learners work with the cards they placed in their desired cards drop area. They are asked to arrange these cards vertically, in order from most to least desired. While canonical Q sort uses a horizontal orientation where the sorting area typically consists of a number of fixed slots where cards can be placed, the present instrument employs a free sort with the idea that the interaction could be made more natural (it was unclear why we should impose a certain structure on the card sort). Text and images are used to help communicate the sorting criterion. Learners complete the activity by dragging

cards from left to right and dropping them in the sorting area. Once dropped, learners are free to reposition the cards as they feel necessary. The sorting area is made wide so that multiple cards can occupy the same vertical position. When half of the cards are sorted, the system displays a feedback message to show appreciation of the learner's effort.

Following the completion of the desiredness sort, the system takes learners through the same activity for the cards they select as undesirable. This part of the activity uses different instructions, appropriate for measuring the extent to which an undesired future is unwanted, but is otherwise equivalent to the desiredness sort.

Sorting by self-efficacy

The activity flow for sorting by self-efficacy is equivalent to above. The system first presents an instruction page with the following content:

Part 2 - Could you make it happen?

Welcome to the second part of the exercise. Here, the focus is on what you feel capable of achieving.

Each of us has different strengths and weaknesses. For example, some of us may be naturally good at learning languages but be shy to start conversations. Others may not learn as easily, but be social and hard working.

Pretend for a moment that you really needed a particular future to come true. For example, imagine that you really want to get a better job. Imagine also that there are many good-paying jobs, but all of them require you to speak Dutch. The question we want you to think about is:

“Could you make it happen?”

If you needed to, could you improve your Dutch to get a better job?

Answer by thinking about your personality, talents, abilities, skills and dedication.

Following the instruction page, the system takes the learner to card sorting activities similar to those used for desiredness, differing only in the use of text and images to communicate the sorting criterion. Learners begin with desired cards first, sorting them according to their perceived self-efficacy, i.e. how capable learners feel of achieving the different futures they selected as desirable.

To avoid exhausting the learners and because the intention is to elicit information concerning futures that individuals feel most strongly towards, and thus could be used as a basis for personalising practice, only the top 15 most-desired cards selected by the learner (obtained in the previous step) are presented for the learner to sort. As with desiredness, following the completion

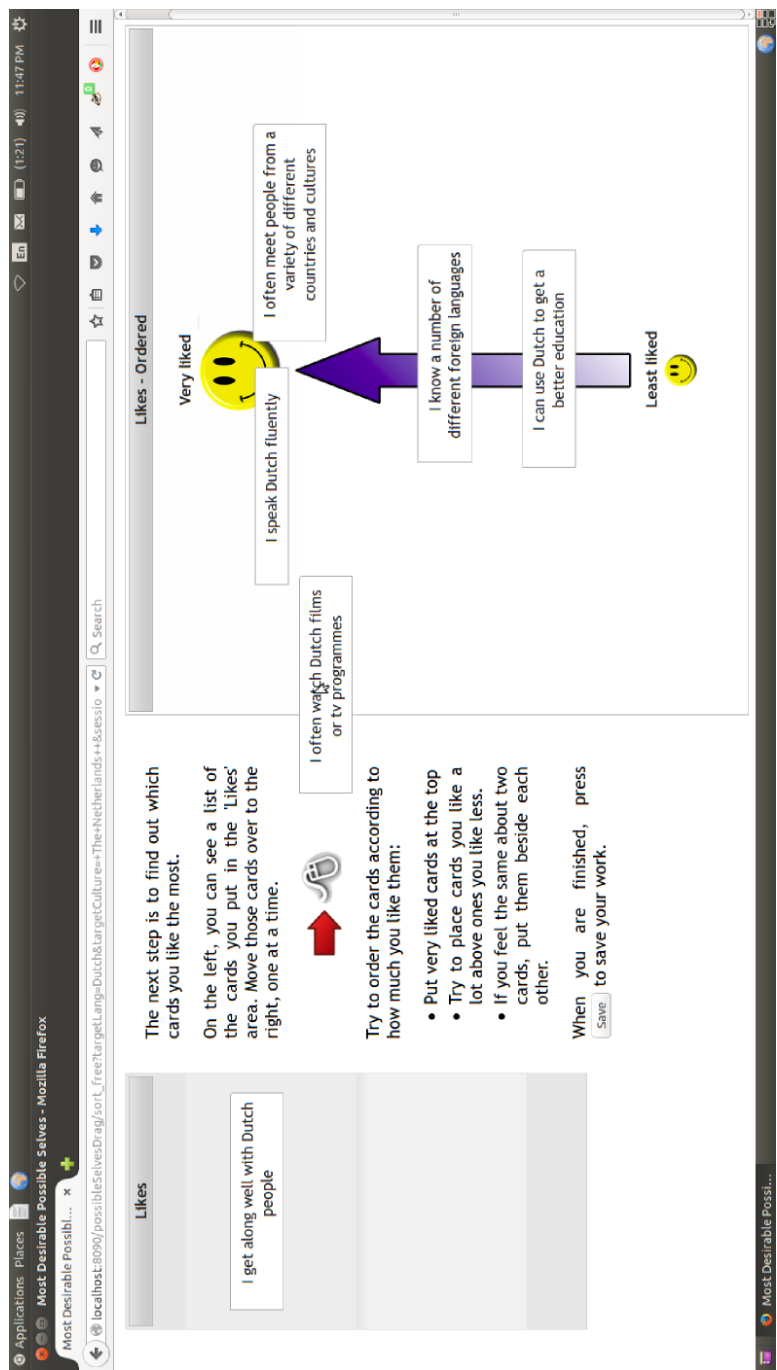


Figure 5.2: A screenshot of the card sorting activity. Learners complete the activity by dragging cards from left to right and dropping them in the sorting area.

of the self-efficacy for the desiredness cards, the system takes learners through the same activity for undesired cards.

Assigning a plausibility rating

The penultimate stage of the activity shifts the focus to learners' beliefs concerning the plausibility of the different futures they have selected (cf. Hessel, 2015). An instructions page presents the following text:

Part 3 - Let's be realistic

Good job, you are almost finished. This is the last part of the activity. There is one last aspect to consider: whether or not it will actually happen. An example: Most of us would probably like the future 'I speak perfect Dutch' quite a lot. However, for those of us who are too busy to study Dutch intensively, this self could be unlikely. Look at the futures below. Realistically, is it likely that they will come true? When you are finished, press 'Next' to continue.

After the instructions, learners are presented with a page displaying the same 15 most-desired possible selves and 15 least-desired selves appearing in the previous stage. For each future self, they are asked to provide a rating on an 11-point scale to indicate how likely it was that the future would actually come true. Here, the futures are presented in semantic differential format to follow the procedure described by Shepard and Marshall (1999) where learners scored cards with points to reflect the plausibility of the future self actually becoming true, and to provide learners with some variety in the activity.

Eliciting personal cards and debrief

Although the activity mainly uses a closed set of items, we were also interested in providing learners with an opportunity to contribute their own personal cards (cf. Irie & Ryan, 2015). After the previous stage, learners are given an opportunity to review the cards they have selected and are invited to contribute personal cards that are not part of their set. Because the main focus is on the card selection and sorting exercises, this part is optional and is incorporated at the end of the activity (Dörnyei, 2003).

Finally, learners are presented with a debriefing activity which thanks them for their participation and informs them of the purpose of the exercise.

5.3.4 Different versions of the activity

During the course of our research, we employed two versions of the card sorting activity to collect data from participants, the original version, as described above in section 5.3.3, and a 'reduced' version. Here, we briefly describe the rationale behind two changes made in the reduced version, and the impact of using two versions during data collection.

The main difference between the original version and the reduced version is that, as the name suggests, the reduced version contains fewer activities. In the reduced version, the card selection part of the activity proceeds as described in section 5.3.3. However, in the steps following card selection, the reduced version omits the card sorting and plausibility rating for the undesired cards.

The rationale for this change is that, over the course of data collection, log file data for the activity and comments from participants indicated that the four card sorting activities and the plausibility ratings in the full version of the activity made significant demands on learners' time. This prompted some reflection on how we might reduce the amount of work in the activity while still meeting our research aims, and on the relative importance of the ideal and ought-to L2 self-guides for the current computer-based learning context.

With respect to research aims, our goal was to explore how L2 self-guides might be used as a resource in computer-based strategy that individualised practice. In this respect, limiting our study to exploring how information on how one of the two self-guide components in the L2MSS could be employed to individualise practice seemed sufficient.

In weighing the two self-guides, we chose to focus on the learners' ideal L2 self, as this information could be treated as a personal goal, and the application of this information as a resource for individualising practice seemed more straightforward relative to the ought-to self guide (which would constitute a kind of 'anti-goal', Carver, 2006). A second reason motivating our decision to focus on ideal L2 selves is that the expected affective benefits of pursuing a desirable future goal are consistent with the selection of constructs that our CALL system tracks during practice (see Chapter 3): learners who perceive practice to be relevant to their goals can be expected to become motivated, while learners who perceive progress towards a desirable goal can be expected to have positive affective experiences (Carver, 2003).

The impact of omitting the card sorting and plausibility activities for undesired selves is obviously that the desiredness, self-efficacy and plausibility data is not generated for undesired futures. However, for the reasons discussed above, we considered this acceptable. Because of the design of the activity which kept sorting activities for the desired and undesired cards separate, it is reasonable to assume that leaving out the sorting and plausibility rating steps for undesired cards did not cause significant differences in card sorting behaviour between the full and reduced versions.

The second difference is concerned with a small change to the user interface for the card selection part and has to do with how learners handle cards perceived to be irrelevant, i.e. those that are neither desired nor undesired. In the full version of the activity (described in section 5.3.3), the system instructs learners to simply ignore irrelevant cards by leaving them where they appear in the card browsing area. In the reduced version of the activity, the card selection of the interface was modified slightly, based on a recommendation from an educator, to include a third card drop area for irrelevant cards, marked with the label 'Not important'. The corresponding instruction page was updated

to reflect this change. We believe this change constitutes a minor usability improvement. However, as both versions of the possible selves activity included a tutorial page providing explicit instructions on how to handle irrelevant cards, from a research point of view, we believe the impact of this difference is negligible.

5.4 Using the activity to investigate the diversity of Dutch learners' ideal L2 selves

As a first step towards investigating the possibility of offering differentiated instruction based on L2 self-guides in our CALL system, we employed the previously described tool to investigate the diversity of ideal L2 selves of a group of Dutch L2 learners who participated in experiments involving the speech-interactive CALL system described in Chapters 3 and 4.

5.4.1 General procedure

Participants for the experiment were recruited from two Dutch language institutes in Nijmegen, the Netherlands. We visited classrooms of Dutch learners at the beginner level (CEF levels A1 - A2) to invite students to participate in an experiment. They were told that they would be working with a CALL system and that they would be completing some exercises from home but that the majority would take place in a university computer lab. Invitations were sent by email, and students who responded were scheduled to come in for practice sessions. Once registered, students received an invitation to complete the possible selves activity from home. Most students completed the activity before coming into the lab, while a small number completed the activity at the lab. Students visited the lab on different occasions to practice their spoken Dutch with the CALL system described in Chapters 3 and 4.

In total, this procedure was run three times, with three different groups having the same range of proficiency levels. In the fall of 2012, 26 learners completed the original version of the possible selves activity and spent two days practising with the system (one day per week, 45 minutes per day). In the spring of 2013, 15 learners completed the original version of the possible selves activity and similarly spent two days practising with the system (also one day per week, 45 minutes per day). In the fall of 2013, 20 learners completed the reduced version of the possible selves activity and spent five days practising with the system (one day per week, 30 minutes per day). Participants also completed a background questionnaire.

5.4.2 Participants

Inspection of the background questionnaire data suggested that two participants did not speak English as a first or second language, but were nevertheless

able to complete the activity. Because of doubts concerning the validity of their data, these participants were removed from further analysis, leaving a total of 59 participants.

Analysis of the background data for the remaining participants revealed that all three groups were culturally and linguistically diverse³. Participants consisted of 23 men and 36 women who had been living in the Netherlands from less than a month to 48 months ($M = 9.5$, $SD = 11.4$). Their age ranged from 19 to 62 years of age, with a mean of 28.5 ($SD = 7.0$). The majority had been living in the Netherlands for less than a year (71.2 %) and had spent a similar amount of time learning Dutch ($M = 5.9$, $SD = 7.7$) with the majority having studied Dutch for less than eight months (78.0 %). The large majority had post-secondary education (50), with the others indicating pre-university education (1), public school (6), elementary school (1) and one answer missing. The land of origin as well as first languages were diverse, with learners from Europe (36), Asia (15), Africa (4) and the Americas (3) and one answer unknown. All learners spoke English as a first or second language. The languages other than their native language and English which they spoke were diverse, with French being the most common (13), followed by German (11) and Spanish (5).

5.4.3 Data preparation

Data from the L2 self-guides activity was exported from the system database and imported into R for analysis. In a first step, the data was plotted to examine how learners used the vertical space during the card sorting activities. During the activity, the vertical sorting space available to learners was adjusted to provide learners with enough space to rank the cards without overlap. On inspecting the data, we observed differences in the way that learners used the space available to them, with the minimum value of the lowest card sometimes much higher than the bottom border of the sorting area. While this could be interpreted as the learners assigning high values to all cards, another plausible explanation is that learners' home computers had monitors with different dimensions, with different sizes of screen immediately available for dragging and dropping (learners were able to scroll lower, but this may be tedious for them given the large number of comparisons and arrangements they had to make). Because of this possibility, we could not assume that the absolute distance from the bottom of the sorting area to the lowest card was meaningful. To work around this limitation, we instead analysed the card sorts relative to their 'active sorting range', i.e. the distance between the top of the sorting area and the learner's lowest card. Card sorts were normalised to the percentage of their height on the active sorting range to enable comparison, with card positions taking on values in the range [5, 100].

³For information on the languages, as well as additional biographical data, see appendix.

5.4.4 The Q analysis

The operationalisation of the ideal L2 self in the present study is different from previous work, which has typically used a single, multi-item scale to obtain one measure representing the general strength of an individual's ideal L2 self (e.g. Papi & Teimouri, 2014; see Hessel, 2015). In the present study, we were interested in whether or not the range of potentially idiosyncratic ideal L2 selves could be represented by a smaller set of *ideal L2 self types*, and if so, how the types might compare in terms of their content.

The approach we chose for examining diversity of ideal L2 selves is based on Q methodology. Q methodology (hereafter referred to as 'Q') is a technique that has been used to study subjective viewpoints among a population, on a broad variety of topics (S. R. Brown, 1980; Dziopa & Ahern, 2011; Irie, 2014; Kerpelman, 2006; Ockwell, 2008; Watts & Stenner, 2005b). Here we will focus on the statistical technique, known as inverted factor analysis, and the qualitative technique used for interpreting the results that is specific to Q. An introduction for language learning researchers has been authored by Irie (2014). S. R. Brown (1980) and Watts and Stenner (2012) have authored comprehensive descriptions of the methodology. For a recent review of literature using Q, see work by Dziopa and Ahern (2011). For recent efforts to incorporate modern advances in statistical techniques into Q, see work by Zabala and Pascual (2016).

The statistical analysis in Q is known as inverted factor analysis. It is 'inverted' in the sense that roles of the variables and the participants are switched. Dimensionality reduction is most commonly performed on a matrix consisting of participants in the rows and variables in the columns, to reduce a large set of variables to a smaller set. In inverted factor analysis, this matrix is transposed so that the rows consist of variables and the columns refer to the participants. In this inverted matrix, card sorts are treated as potential dimensions of learner space. A matrix of inter-participant correlations is then calculated and submitted to principal component analysis (PCA). The output of PCA is a smaller number of orthogonal vectors, called *factors*, that capture main directions of variance, and which through linear combination, can be used to construct the original card sorts. The association of a card sort with a factor is measured by degree of loading. Card sorts which load significantly are associated with that factor. The final step of the statistical analysis is to compute so-called *factor arrays*. A factor array is computed by taking the card sorts which load significantly onto a factor and calculating the weighted mean based on the loading coefficient for each factor (see Zabala, 2014). This procedure produces a hypothetical card sort representing the weighted contributions of the individual learners' card sorts. In the context of the present study, factor arrays represent hypothetical ideal L2 self types.

Note that, as factor loading scores are the mechanism used to associate card sorts with hypothetical types, it is possible that card sorts can be related to the types in several, not always straightforward ways. In the simplest case,

card sorts load positively and significantly onto one and only one factor. In this case, learners can be interpreted to have an ideal L2 self type that is similar to the hypothetical card sort (i.e. factor array) for that factor.

It is also possible that a card sort loads significantly onto multiple factors, which would suggest that the associated learner has an ideal L2 self that is similar to two or more ideal L2 self types. In the literature, card sorts that load onto multiple factors are known as *confounded card sorts* (Dziopa & Ahern, 2011; Zabala, 2014), and are usually not interpreted (Akhtar-Danesh, Baumann, & Cordingley, 2008). However, Borgen and Barnett (1971) suggest that “if the objective is an exploratory study of the structure of a set of data rather than the construction of an empirical classification the split factor loadings may have some utility” (p. 586). In the present study, confounding card sorts are interpreted to represent ideal L2 selves that have similarities with multiple hypothetical ideal L2 self types (see below).

A third and final possibility is that a card sort loads onto a factor negatively (Blashfield & Aldenderfer, 1988). In this case, the learner can be interpreted to have an ideal L2 self that is similar to the *inverse* or mirror image of the hypothetical ideal L2 self type. According to Watts and Stenner (2012), interpretation of card sorts that load negatively is typically done in one of two ways: either the factor array is reversed, so that high ranking cards take on low positions and vice-versa, and this is interpreted, or a separate factor array is calculated using only the participants who load negatively. In the present study, two or fewer card sorts loaded negatively onto each bipolar factor. Watts and Stenner (2005a) seem to consider it necessary to interpret negative factors regardless of how many card sorts actually load negatively. However, Dunnette, Campbell, and Hakel (1967) find that it does “not seem useful descriptively to attempt a separate interpretation of just one or so few persons’ responses”, noting however that “the occurrence of such “singletons” in our cluster analyses is further testimony to the prevalence of marked individual differences” (p. 164). In the present study, we follow Dunnette et al. (1967): we do not provide an interpretation of hypothetical ideal L2 self types that result from card sorts loading negatively onto factors, as too few participants load onto these factors to consider them a representative type in our data. However, we do regard these card sorts as distinct ideal L2 self-guides and consider them in our analysis of the prevalence of the types.

To perform the inverted factor analysis on our card sort data, we used the R package *qmethod* (Zabala, 2014), modified slightly to accept the continuous card position values generated by our sorting activity. Before submitting the card sorts to an inverted factor analysis, we had to specify the number of components to extract (Zabala, 2014). A variety of criteria is available for selecting the number of factors, including scree plots, the eigenvalue > 1 rule (S. R. Brown, 1980) and parallel analysis. In our analysis we used the *fa.parallel* function from the *psych* R package which computes the parallel analysis technique. The technique compares the scree plot for the recorded data against a scree plot obtained from a principal component analysis of

a set of random (and therefore factorless) data as a benchmark. From this analysis, the scree plot suggested 3 components, the eigenvalues 1 > rule produced 15 components, while the parallel analysis output varied between 2 and 3 components. We compared the 2- and 3-factor solutions, obtained via qmethod's inverted factor analysis using varimax rotation, choosing the 3-factor solution on the basis that it accounts for more total variation (Zabala, 2014), more card sorts load onto the factors significantly (S. R. Brown, 1980) and its interpretation appeared to connect better with the literature (Watts & Stenner, 2012).

Participant loadings onto the three components are shown in the appendix (Table 5.4). A card sort is considered associated with a component if two criteria are met: 1) the loading is significantly high ($p < .05$) and 2) the square of the factor loading is larger than the sum of the squared factor loadings for all other factors (Zabala, 2014). The threshold for significance is given by the formula $1.96 (1 / \sqrt{N})$, where N is the number of cards (Zabala, 2014). In the present study, a loading reached significance at .37.

Ideal L2 self types

To analyse the hypothetical card sorts that emerge from Q analysis, Irie (2014) recommends using crib sheets, as described by (Watts & Stenner, 2012), for each factor that emerges from inverse factor analysis. This involves looking at the two extremes of the sorting range, and also involves considering ratings of statements which are either significantly lower or higher than those given by other groups. Canonical Q Sort uses a fixed distribution for their sorting area in which the extremes are the outermost two sorting columns consisting of five cards per side. We adapted this technique in our analysis of the data by considering the top and bottom five cards in a card sort as extreme cards representing a learner's most desired future selves at the top and, at the bottom, future selves towards which the learner felt most indifferent. When considering whether two ratings for the same card are significantly different, we use the facility included in qmethod (Zabala, 2014) which calculates a threshold based on the standard error of differences⁴. Following Q methodology we draw on these notes and qmethod's table of distinguishing and consensus statements (see Table 5.1) to produce a textual interpretation of each hypothetical card sort, representing an ideal L2 self type.

Prevalence of the types

To obtain information on the diversity of ideal L2 selves in our population, we calculated the number of distinct future self configurations, using the factor loading table (Table 5.4 in the appendix). This table provides information on the association between a card sort and each of the hypothetical card sorts which emerged from the inverted factor analysis. If a card sort loads

⁴2.58 * SED for $p < .01$, 1.96 * SED for $p < .05$

significantly onto one and only one of the hypothetical card sorts following the two criteria outlined above, they are counted as an instance of that ideal L2 self type. In the case where a factor is bipolar, i.e. there is at least one card sort which loads negatively onto a hypothetical card sort, the card sort is assumed to represent a distinct perspective (Watts & Stenner, 2012), because L2 practice exercises tailored to the motivational configuration of individuals who positively associate with an ideal L2 self type could not be expected to be suitable for learners who are negatively associated with that type. Therefore, learners loading negatively loading onto a factor are counted as an instance of a distinct ideal L2 self type.

As discussed above, it is also possible to have confounded card sorts. These confounded sorts contribute information related to the diversity of ideal L2 selves: In the case where a card sort loads significantly onto multiple perspectives, we interpreted the card sort to share the ideal-self characteristics of the associated factors, and therefore the card sort is not treated as a distinct perspective. These card sorts are counted towards all of the types with which they are associated. Finally, card sorts which do not load significantly onto any factor are treated as a distinct perspective.

Self-efficacy and plausibility associated with the types

Due to the design of the activity, which asked learners to provide self-efficacy and plausibility ratings for only their 15-most desired cards, our approach described above for analysing learners' ideal L2 selves along the dimension of desiredness was necessarily different from how we approached self-efficacy and plausibility. The approach we used for the latter two dimensions can be described as a 'second-tier' examination which looked at the extent to which individuals associated with a particular desiredness group feel that they have the capacity to achieve their ideal L2 self and the extent to which they perceive that self to be realistically attainable.

We divided the card sorts into subsets corresponding to their ideal L2 self types produced from the inverted factor analysis. This step partitioned the learners into groups with similar most-desired future selves. We then calculated the mean self-efficacy and plausibility for each individual based on their top N most-desired possible selves. Here, N was chosen to be four, rather than five, due to a technical error in the fall 2012 experiment which prevented some users from rating all of their top 15 most-desired selves. This procedure yielded one row per individual, each row constituting an approximate measure of each individual's self-efficacy and plausibility beliefs towards their ideal L2 self. To investigate whether learners could be classified into sub-types based on either of these dimensions, we both visually inspected the distributions and applied a statistical test (Hartigan & Hartigan, 1985) to determine whether the distribution of self-efficacy and plausibility levels among the members of the group was unimodal. If the distribution was found to be multimodal, this could be an indication that there is more than one central tendency among learners

associated with a particular ideal L2 self type, which would in turn suggest the presence of sub-types.

5.5 Results

We first analyse the results of the inverted factor analysis and the contents of the ideal L2 self types. Due to space limitations, only types with three or more loadings are described. Following that, we look at the prevalence of the types in our population of learners, before examining the self-efficacy and plausibility characteristics of the most-desired selves in the types.

Statistical comparisons of the three profiles produced in the inverted factor analysis revealed a number of significant differences on the desiredness ratings of the 28 future selves. These differences are detailed in Table 5.1 and presented visually in Figure 5.3.

5.5.1 Contents of the ideal L2 self types

Following Q methodology (S. R. Brown, 1980; Watts & Stenner, 2012), we interpret the factor scores (i.e. ideal L2 self configurations) visible in Figure 5.3 with a textual description. In the text, parentheses with two values are used to refer to a future self item by its number, and the value of the item in the normalised factor array rounded to the nearest integer. For example, (3: 76) indicates that item 3 has a vertical position at 76 % of the height of sorting area.

Type 1+ - Functional and pragmatic without social or cultural interests

The future self configuration represented by Type 1 appears to be most concerned with themes related to autonomy, Dutch proficiency and the pragmatic benefits of learning Dutch. Type 1 possesses a strong desire to communicate with native speakers without help (2: 91) and to function on a day-to-day basis without reverting to English or their native language (21: 86), at a level significantly higher than the other groups. At the same time, they also view the ability to speak Dutch as highly valuable, with a strong desire to feel comfortable while speaking Dutch (4: 82), a characteristic that distinguishes them from Type 2 but not Type 3. The importance Type 1 places on speaking Dutch fluently (11: 79), having the capacity to express their emotions (16: 78), and feeling at ease while speaking in a group (14: 77) distinguishes them from the other types. Together, this suggests that spoken Dutch is an important aspect of their envisioned daily lives, although they do not place as much importance on speaking Dutch fluently as Type 3.

Table 5.1: Output of the qmethod table showing similarities and differences between the types. Under ‘Outcomes’, ‘Dist.’ refers to distinguishing features, indicating that a types’ value is significantly different from the ratings of one or more type(s). Similarly, ‘Consensus’ indicates all types have equivalent values. T1-T2 is the difference between desiredness rating for Type 1 and Type 2. Units are z-scores. The text for item 21 is completed with ‘speaking English or my native language’.

** $p < .01$, * $< .05$.

Item	Outcome	T1-T2	sig	T1-T3	sig	T2-T3	sig
I have made people in my life happy or proud by learning Dutch (1)	Dist. T1 only	0.65	**	0.8	**	0.15	
I can communicate with native speakers in Dutch without the help of others (2)	Dist. all	0.74	**	1.29	**	0.55	**
I feel comfortable when I speak Dutch (4)		0.5	**	0.19		-0.31	
I have people in my life who think it’s great that I study Dutch (6)	Dist. T1 only	1.01	**	0.92	**	-0.09	
I feel confident enough to begin conversations in Dutch (9)	Dist. T3 only	-0.24		-0.96	**	-0.72	**
I can write fluently in Dutch (10)	Dist. all	2	**	-0.41	*	-2.41	**
I speak Dutch fluently (11)	Dist. all	1.16	**	-0.86	**	-2.02	**
I feel comfortable speaking Dutch in a group (of 5 or more people) (14)	Dist. T1 only	0.74	**	0.82	**	0.08	
I can express my emotions in Dutch (e.g. happiness, anger, sad, love) (16)	Dist. T1 only	0.81	**	0.46	*	-0.34	
I can understand Dutch slang and use it properly (17)	Dist. all	0.92	**	-0.39	*	-1.32	**
I can express myself creatively in Dutch (20)	Dist. T3 only	0.3		-0.93	**	-1.23	**
I can function in my day-to-day life in the Netherlands without ... (21)	Dist. all	0.54	**	3.26	**	2.71	**

I can use Dutch to get a better job (22)	Dist. all	1.29	**	0.69	**	-0.61	**
I speak a foreign language often (24)	Dist. all	-2.64	**	-0.48	**	2.17	**
I know a number of different foreign languages (25)	Dist. T2 only	-2.18	**	0.22		2.4	**
I can use Dutch to get a better education (26)	Dist. T2 only	1	**	-0.04		-1.04	**
I often meet people from a variety of different countries and cultures (30)	Dist. all	-2.86	**	-1.19	**	1.67	**
I have opportunities to learn about the values and customs of (35)	Dist. all	-2.12	**	-0.58	**	1.53	**
I get along well with Dutch people (36)	Dist. all	-1.51	**	-1.09	**	0.42	*
I have opportunities to meet Dutch people (37)	Dist. T1 only	-0.88	**	-0.83	**	0.06	
I spend a lot of my time with Dutch people (38)	Consensus	-0.15		0.01		0.16	
I have opportunities to spend time with other Dutch learners (39)	Dist. T1 only	-0.47	**	-0.75	**	-0.28	
I often watch Dutch films or tv programmes (42)	Dist. T2 only	0.58	**	0.07		-0.51	**
I often read Dutch books, magazines, or newspapers (43)	Consensus	0.24		-0.02		-0.26	
I am similar to Dutch people (44)		0.36	*	0.13		-0.24	
I have good friends who are Dutch (46)	Dist. T1 only	-1.08	**	-0.98	**	0.1	
I live in the Netherlands permanently (47)	Dist. T2 only	1.17	**	0.33		-0.84	**
I understand Dutch people and their way of life (48)	Consensus	0.09		0.31		0.22	

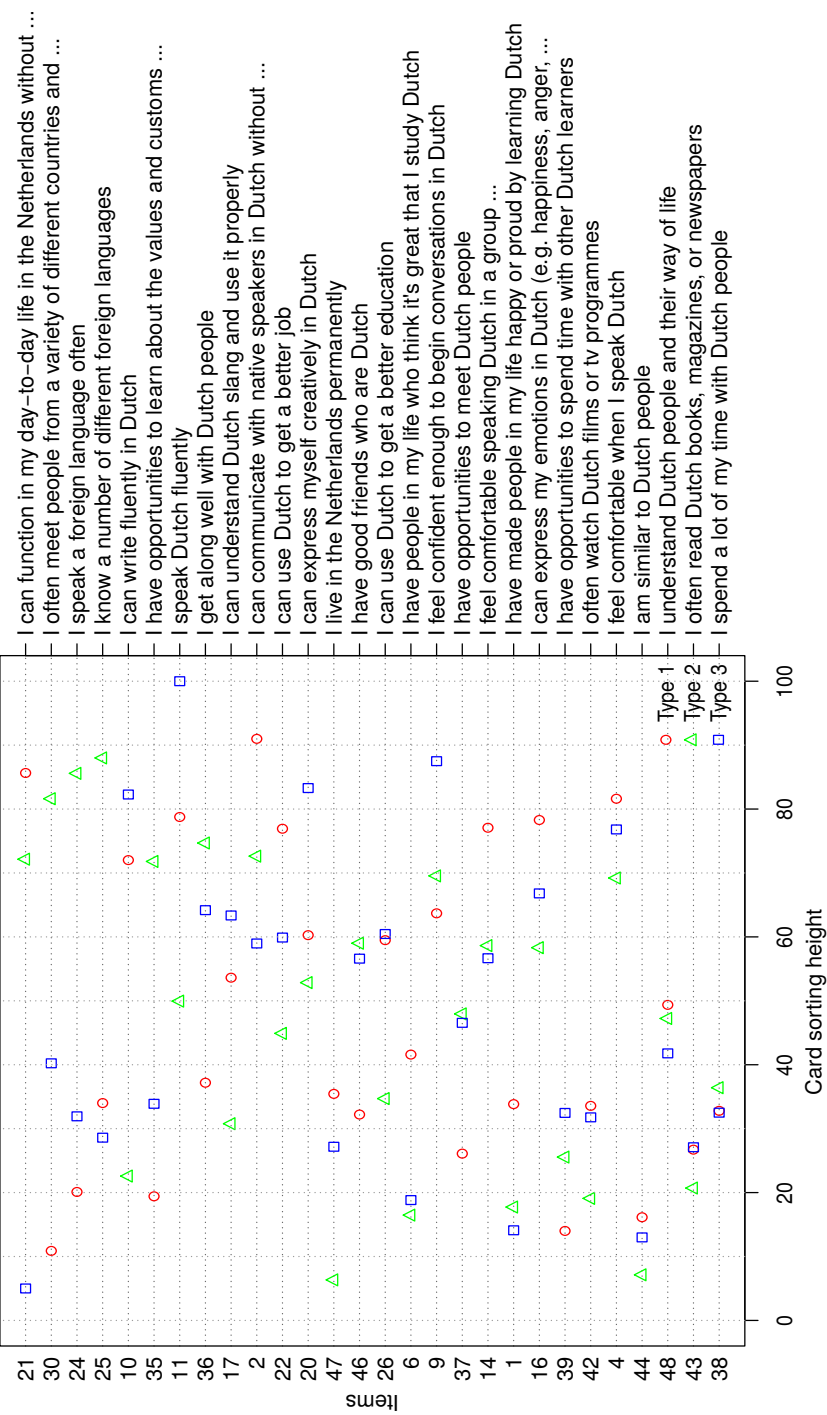


Figure 5.3: A qmethod plot showing the hypothetical card sorts for the three types emerging from the analysis. For convenience, scores have been scaled to the range [5, 100] to match how they would have appeared in the card sorting activity.

Concerning pragmatic benefits such as using Dutch to get a better job, Type 1 attaches significantly more importance compared to other groups (22: 77). Their views regarding potential educational benefits can be characterised as mid-range (26: 59), and are similar to Type 3. Along with expressing their emotions, Type 1 is distinguished from the other groups by their moderate desires to write fluently in Dutch (10: 72) and be familiar with Dutch slang (17: 54), qualities which are likely to facilitate their functioning in daily life. Seen within their profile, Type 1's concern about pleasing others in the future by learning Dutch appear moderate. However, comparing values with other types for futures in which they are appreciated for studying Dutch (6: 41) and in which they have made others happy or proud by learning Dutch (1: 33) reveals that Type 1+ is in fact significantly more concerned about this aspect of their future.

In examining the cards at the lower end of the sort, we see that one striking characteristic of Type 1 is their seeming lack of interest in the potential social or cultural benefits of learning Dutch, or of learning foreign languages in general. Learners with Type 1 perspective appear unmoved by the possibility of becoming closer to Dutch people, evidenced by the relative lack of importance they place on getting along well with Dutch people (36: 37), having good Dutch friends (46: 32), and on having opportunities to meet Dutch people (37: 26), all selves that are significantly less desirable in this perspective. Type 1 also appears to lack a general interest in foreign languages and cultures. They are indifferent to a future in which they speak a foreign language often (24: 20). They also appear uninterested in learning about the values and customs of other cultures (35: 19) and are unmoved by the potential of having opportunities to meet people from other countries and cultures (30: 11). They similarly place a low priority on having opportunities to spend time with other Dutch learners (39: 14). Regarding these four possible futures, Type 1 reports significantly less desiredness compared to the other groups. Finally, although Type 1 is the most positive concerning the possibility of becoming similar to Dutch people (44: 16), significantly different from Type 2, like the other groups (see below), this future is one of the least attractive.

Type 2+ - Social L2 enthusiast lacking strong Dutch-specific interest

A Type 2 perspective on the ideal L2 self reflects a strong interest in speaking foreign languages and potential social benefits of speaking an L2. Type 2 represents the individual who aspires towards a future in which they know a number of different foreign languages (25: 88) and speak those languages often (24: 86) characteristics that are significantly higher in this perspective. Learners with a Type 2 perspective also appear keen to have contact with people from different countries and cultures (30: 82), a characteristic that distinguishes them from the other types. Given the above and the fact that they were living in the Netherlands at the time, it seems logical that they report a strong desire for a future in which they get along well with Dutch people (36: 77),

significantly higher than the other types. From the perspective of Type 2, a future in which they are able to communicate in Dutch with native speakers without assistance (2: 73) and without resorting to speaking English or their native language (21: 72) is highly desirable and a distinguishing feature, with these aspects valued more than Type 3 yet less than Type 1. Of the three types, Type 2 attaches the most importance to learning about foreign cultures (35: 72). Together these characteristics seem to suggest that learners with a Type 2 perspective have a keen interest in interacting with people from around the world and see foreign languages as a means to making that happen. Consistent with this interpretation are the positive, but moderate ratings for futures in which they have good Dutch friends (46: 59) and opportunities to meet Dutch people (37: 48), with these aspects reaching levels similar to that for Type 3.

Turning to characteristics of the future which hold less appeal, we seem to observe a pattern suggesting that learning Dutch may only have a moderate role in the future. First, this perspective is relatively indifferent towards becoming a fluent Dutch speaker (11: 50), with a desiredness rating significantly lower than the other types. Other futures referring to more specific capacities with the Dutch language, namely the ability to use Dutch slang (17: 31) and to write fluently (10: 23), are even less moving and significantly lower than in other perspectives. Second, Type 2 has distinctly moderate views towards pragmatic benefits of learning Dutch, such as improving one's career (22: 45) or education (26: 35), which are in fact also significantly lower than those of the other perspectives. While watching Dutch television or film was not a highly desired characteristic of any of the perspectives in this study, Type 2 is significantly more indifferent to this aspect of their future (42: 19). Learners with a Type 2 perspective also attach little importance to pleasing others by learning Dutch (1: 18) or being appreciated for learning Dutch (6: 17), a characteristic shared with Type 3 but significantly different from Type 1. Finally, Type 2 is the most indifferent to the prospect of becoming similar to Dutch people (44: 7), distinguishing them from Type 1, but unlike Types 1 and 3, they are also unmoved by the possibility of living in the Netherlands permanently (47: 6).

Type 3 - Aspiring for proficiency yet acknowledging other potential benefits

A Type 3 configuration of an ideal L2 self can be characterised as emphasising aspects of the future related to becoming proficient in the Dutch language. Type 3 attaches the highest importance to becoming a fluent Dutch speaker (11: 100), a characteristic rated significantly higher than the other groups. In addition to becoming fluent, this perspective expresses strong hopes of feeling confident enough to begin conversations (9: 88) to the degree that this distinguishes them from other perspectives. Type 3's strong aspirations for Dutch proficiency extend past speaking to express a desire for writing fluently (10: 82), having the capacity to express themselves creatively (20: 83), and using and understanding Dutch slang (17: 63), all rated significantly higher than the

other groups. In addition to being proficient with the Dutch language, this perspective also hopes to feel at ease while speaking Dutch (4: 77), to an extent comparable to the other groups. Although proficiency is their priority, they attach moderate importance to practical benefits. They look favourably towards the possibility of improving their career by knowing Dutch (22: 60), with this aspect significantly higher than Type 2 and less than Type 1. With regards to possible educational benefits, they hold positive views (26: 61) at a level similar to Type 1. However, with respect to communicating in Dutch with native speakers without help, Type 3 attaches significantly less importance (2: 59) relative to the other groups. Possible selves related to potential social benefits appear to have moderate to somewhat low importance, with Dutch-specific selves appearing more important. Type 3 expresses mid-range hopes to get along well with Dutch people (36: 64), at a level significantly higher than Type 1, but less than Type 2. The related futures involving having good Dutch friends (46: 57) and opportunities to meet Dutch people (37: 47) also have moderate importance, at a level significantly higher than Type 1, but comparable to Type 2.

Regarding characteristics of the future viewed as less important by Type 3, we find that aspects of their future related to learning about foreign cultures and languages in general attract only moderate interest. They appear somewhat interested in meeting people from abroad (30: 40) and learning about foreign cultures (35: 34), with values for these futures significantly higher than Type 1, yet lower than Type 2, though the possibility of understanding the Dutch way of life is seen as slightly more desirable (48: 42), the value for this future being comparable to the other groups. Although Type 3 appears very interested in learning Dutch, the futures related to speaking a foreign language often (24: 32) and knowing a number of foreign languages (25: 29) have lower priority, although the former is important enough to be distinct from the Type 1 perspective. Type 3 learners do not have a particularly strong desire to live in the Netherlands permanently (47: 27), although they value the possibility significantly more than Type 2. While this perspective appears to value becoming a highly proficient Dutch speaker, like the other types, it is unmoved by the prospect of reading in Dutch (43: 27). Neither does this type attach importance to pleasing others by, or being appreciated for learning Dutch (respectively, 1: 14, 6: 19), a characteristic they share with Type 2 learners. Like the other perspectives, becoming similar to Dutch people appears irrelevant to their future concerns (44: 13). Interestingly, this perspective attaches the least amount of value to being able to communicate without using English or their native language (21: 5), much less than groups 1 and 2. This characteristic, when combined with the lack of importance placed on speaking foreign languages often, suggests that while Type 3 learners are very motivated to learn Dutch, becoming fully functional or immersed in Dutch on a day-to-day basis is not an important aspect of their future.

5.5.2 Prevalence of the types

Inspection of the loading table produced by the inverted factor analysis (see Table 5.4) shows that factors 1 and 2 have participants who load positively and negatively. As these express opposing views (Watts & Stenner, 2012), each should be interpreted as two distinct perspectives. Participants associated with Type 3 load positively onto the component, contributing one perspective.

In addition, we have 12 individuals who failed to load onto one and only one of the hypothetical perspectives (see Table 5.4 in the appendix). Of these, we observe that two participants could not be classified because they loaded significantly onto more than one component: one participant loaded significantly onto the components representing Type 2 and Type 3, while the other loaded significantly onto Type 1 and Type 3 components.

A third participant loaded significantly (and negatively) onto Type 3, but the loading value did not satisfy the second, squared loading criterion (see section 5.4.4). This participant contributes a unique perspective.

The remaining nine individuals did not load significantly onto any of the components and each contribute a distinct perspective. The total number of unique perspectives in this group of learners is 15.

Table 5.2: Prevalence of the ideal L2 self types among the 59 participants. The column ‘single-type’ reports the number of learners who have these ideal self characteristics and who loaded onto that type and only that type, while the ‘multiple-type’ column reports the number of learners with these characteristics originating from learners who loaded onto multiple self types.

	Single-type	Multiple-type	Prevalence (%)
Type 1+	16	1	17 (28.8)
Type 1-	2	-	2 (3.4)
Type 2+	15	1	16 (27.1)
Type 2-	1	-	1 (1.7)
Type 3	13	2	15 (25.4)
Other	-	-	10 (16.9)

Table 5.2 shows the prevalence of the types, where we can observe that the three major types were more or less equally present. Type 1+, a future-perspective emphasising functional and pragmatic aspects of learning Dutch without social or cultural interests, was held by 17 individuals; Type 2+, an ideal self prioritising social aspects of learning an L2 but lacking strong Dutch-specific interest, was held by 16 individuals, and Type 3, associated with a focus on becoming proficient in the Dutch language but acknowledging other potential benefits, was held by 15 individuals. The Type 1- (the group with members who load negatively onto component 1) has two members, while the remaining types have 1 member each.

5.5.3 Self-efficacy and plausibility associated with the types

Visual inspection of the data for the participants' mean self-efficacy and plausibility for their most-desired selves suggested that the distributions had a single central tendency, indicating a lack of subgroups within an ideal L2 self type. Table 5.3 shows that this was supported by Hartigan's dip test. Although the contents of learners' desired selves differed, their beliefs concerning the self-efficacy and plausibility of the most desired facets of their ideal L2 self-guides seem to conform to one type.

Table 5.3: Results of the dip test for unimodality - multimodality.

	Self-efficacy		Plausibility	
	D	p	D	p
Type 1+	0.09	.29	0.10	.15
Type 2+	0.06	.93	0.08	.69
Type 3	0.10	.61	0.10	.70

5.6 General Discussion

The present study drew on developments in the L2 motivation and broader psychology literature to develop a computer-based method for eliciting information about language learners' future selves and investigate how such a computer application could be useful for personalising language instruction. It searched for motivational types for a population of learners learning a language other than English and for a range of ages rarely addressed in previous studies (adult L2 learners), using a methodology that analysed facets of learners' ideal L2 self-guides holistically. In this section we reflect on the work completed by reviewing our research questions.

5.6.1 What might a computerised tool that elicited information on learners' L2 self-guides look like?

In considering what method to use to elicit data on learners' self-concept, researchers have to think about their own use case and select methods which best meets their needs (Packard & Conway, 2006). The goal of the present study was to develop an activity that obtains information on future selves for the purpose of providing motivationally relevant practice. We also had the requirement that the activity use technology to reduce the amount of work necessary to run the activity, by automating as much of it as possible. This would make it more scalable and feasible for language educators to use and also support embedding it as a data source in a tutorial CALL system. We also wanted to contribute a method that exploited modern web technologies

to develop an activity that was engaging for learners. The characteristics of card sorting seemed to make it well suited for our application: card sorting is thought to provide more opportunities for participants to interact with the items (Packard & Conway, 2006); learners have been reported to enjoy the activity (Pemberton & Cook, 2012); the method is structured enough to be implemented as a stand-alone computer application, and the data it elicits is readily processed by a computer.

The computer application we have developed provides one answer to how technology can be used to create a reflective activity that elicits data for the purpose of personalising instruction. The activity uses web technology to provide an interactive card sorting activity based on the Possible Selves Mapping Interview technique (Shepard & Marshall, 1999). The activity elicits properties of an individual's L2 self-guides that are thought to be relevant for learning (i.e. desirability, self-efficacy, plausibility). It also aims to provide a tutorial-style interaction by including instructional pages with texts designed to stimulate the learner to reflect on different possible futures. By utilising modern web technologies, we were able to provide an interface that allows learners to sort the cards freely. A logging facility made free card sorts possible without creating extra effort, as records of the height of the sorting areas and position of the cards allowed each participant's sorts to be examined. These data could, in turn, could be used in a Q analysis. To the best of our knowledge, this is the first computer-based card sorting activity for L2 self-guides to be developed.

Having discussed the application's positive characteristics, we can also share some of its limitations. First, a review of log files for the activity and participants' comments concerning the amount of work involved suggested that the original version of the activity, which included steps for surveying characteristics of learners' desired and undesired futures, was too long to complete in one sitting. To address this issue, we created a reduced version of the activity that did not elicit data on learners' undesired futures.

However, for research interested in studying self-based approaches to motivation, this strategy of reducing the number of items is not an optimal solution. This is because there are other selves that would be interesting from a theoretical point of view. For example, undesired selves are likely to be related to Dörnyei's ought-to L2 self, which is a main component of the L2MSS and theorised to motivate learner behaviour to prevent certain undesired futures. Previous work investigating the motivational types of language learners has found that the strength of an ought-to self differs with individuals, and that it is a characteristic useful when considering motivational types (e.g. Papi & Teimouri, 2014). To mention a second example, the application as it is now does not elicit information on the learners' present self-concept, while we have seen that this is an important aspect of self-based approaches to L2 motivation. In future studies investigating the conditions under which a future self becomes motivating, it would be very important to obtain data on the present self-concept, in order to be able to detect areas of self-discrepancy. Similarly, we may also want to include additional sorting activities to explore

which dimensions are related to a possible selves capacity to motivate (Hessel, 2015). One solution that might help to avoid overloading the learner is to add a save feature, to allow users to take breaks resume their work later. This could be combined with an experiment design that spread the activity over a number of days.

A separate issue has to do with the continuum displayed in the activity's card sorting areas. In canonical Q sorting, it is standard for learners to sort along a bipolar continuum, such as 'like me' and 'not like me' or 'feared' and 'hoped for', where the ends represent opposite poles of a continuum. In the present version of the activity, the two ends represent indifference, at the bottom, and strong positive magnitude, at the top. To better align the activity with standard Q methodology practices, we might investigate using a bipolar sorting area, where the middle represents indifference, and combine sorting desired and undesired futures into one step.

A final point of discussion relates to the item inventory used in the application. While these items represent important motivational constructs and are likely to be an approach appropriate for integrating the activity into a tutorial CALL system, for research purposes it could also be interesting to have learners generate their own cards, to hear about their ideal L2 selves in their own words. Currently, the application presents the card generation step at the end of the activity, but it would be trivial to move that stage to the beginning of the activity. Configured this way, learners could then generate cards and in later stages sort them. While this could introduce difficulties in comparing ideal self types, one way to overcome this could be to employ qualitative analysis techniques to group cards with similar themes. Another potential idea would be to organise an initial phase where learners create their cards, and have all cards go to a common pool, after which they are used in the selection and sorting activities.

5.6.2 What motivational types would emerge from application of this tool to language learners?

Q analysis suggested the presence of three major ideal L2 self types. The Type 1+ ideal self prioritises becoming competent, feeling at ease while speaking Dutch, and speaking fluently, and pragmatic aspects such as being able to get a job, aspects which are likely to make one's quality of day-to-day life better while living in the Netherlands. At the same time, learners with this perspective appear unmoved by the possibility of becoming closer to the Dutch community and learning about its culture. Taking these attributes together, we might characterise them as valuing futures in which they are linguistically competent, confident and speak fluently; they have the highest instrumental orientation relative to the other groups, and lack in positive regard to the L2 community, integrative orientation, general interest in foreign languages and international posture.

In their study, Csizér and Dörnyei (2005a) found a group with similar

characteristics to Type 1+. Their group was found to have “a powerful sense of the professional relevance of L2 proficiency, but [that] this is not accompanied by a similarly positive regard toward the L2 culture and community” (p. 638). Similarly, in the other cluster analysis study investigating motivational types, Papi and Teimouri (2014) write with respect to learners in their Group 3 that they “have shaped the professional aspect of their ideal L2 self, are concerned about the negative consequences of failure, and do not hold positive attitudes toward the L2 community and its culture” (p. 513). Thus, the Type 1+ ideal self found in this study seem to correspond well with previous research.

The Type 2+ ideal self values a future in which one has learned a number of foreign languages and has opportunities to speak those languages often, likely while meeting people with a variety of backgrounds from different countries around the world. With this interest comes, as one might expect, a strong desire to get along well with the people of the Netherlands and to be able to communicate with them. Concerning the language, the emphasis seems to be on becoming competent and functional in Dutch, rather than on highly proficient. Furthermore, this perspective attaches relatively little importance towards the possibility of improving their career or education, and appears to be indifferent to living permanently in the Netherlands. In terms of L2 motivational constructs, this group can be characterised as having primarily a strong interest in foreign languages and high values for international posture, as well as a desire to be linguistically competent. They appear to have moderate integrative orientation, a weak instrumental orientation and a low priority for living in a milieu that appreciates their efforts to learn Dutch.

The closest match to Type 2+ in the study by Csizér and Dörnyei (2005a) is Group 2, described as having positive attitudes towards the L2, but not seeing it as “related to their future professional life ... They simply do not think that they will need the L2” (p. 638). In their cluster analysis, Papi and Teimouri (2014) finds a similar group, consisting of members that “have not developed strong L2 selves but hold positive attitudes toward the target community and culture and seem to be motivated mainly by the executive motives in their language learning environment” (p. 513). Like these groups, the Type 2+ perspective also has moderately positive regard for Dutch people and culture, yet, while interested generally, perhaps because they were living in the Netherlands at the time, they do not seem to find Dutch highly necessary for their future.

The third major type in our study was Type 3, which can be characterised as having a strong interest in becoming proficient with the Dutch language. The ability to speak fluently and confidently ranks highest with this perspective, yet interest in the Dutch language also extends to other aspects, such as writing fluently. Although it prioritises the language, this ideal self also has the characteristics of viewing potential pragmatic and social benefits of learning Dutch somewhat positively. Relative to learning Dutch, the prospect of learning other foreign languages appears uninteresting to this perspective, as do futures related to pleasing others by learning Dutch. In terms of

L2 motivational constructs, this perspective values linguistic proficiency, and becoming a person who is highly willing to communicate in Dutch. They have moderate instrumental orientation, integrative orientation, attitudes towards native Dutch speakers, a relatively low interest in foreign languages and are indifferent to living in a social milieu that values their Dutch learning.

Type 3 also seems to connect with previous work. In the study by Csizér and Dörnyei (2005a), this type would appear to match Group 4, the group who had high levels of motivation on variables related to the L2 culture and community as well as variables related to professional and educational development. There also appears to be a match with the work of Papi and Teimouri (2014), in which two groups were positive towards a variety of motivational dimensions. They are open to learning about other cultures and languages, but also have strong pragmatic motivations. According to Papi and Teimouri (2014), the difference between these groups has to do with the strengths of each group's ought-to self: Group 4 was found to have a weak ought-to self and be less motivated by obligations relative to group 5. Unfortunately, it is difficult to say whether Type 3 in our study corresponds better with group 4 or group 5. This is because we did not include an analysis of learners' undesired selves (i.e. future selves learners might feel obligated to prevent). However, we observe that learners associated with the Type 3 ideal self seem to be indifferent to futures in which they please others, which tentatively suggests that their ought-to self-guide might not have a strong role, potentially making Type 3 more similar to Group 4.

Another finding of interest is provided by Irie and Ryan (2015): the inverted factor analysis employed by the study to analyse changes in learners' L2-related self concept after a study abroad experience found three distinct viewpoints, with the "comfortable users" viewpoint being most relevant here. That view of language learning reflects a person who had become comfortable with learning and with L2 use, who had "made friends and enjoys communicating with them in the L2" and who "still believes in the possibility of achieving a high level of proficiency" (p. 355). This description seems to correspond with the future-perspective of Type 3; we might say that Type 3 learners hope to one day become the person represented by the "comfortable user" self-concept.

Finally, there is at least one type observed in previous work that was not present here. As discussed above already, the activity did not support an analysis of learners' undesired future selves. This prevented the possibility of detecting types motivated primarily by an ought-to L2 self, such as the duty-bound learner reported in the study by Irie and Ryan (2015), which was interpreted to lack a strong ideal L2 self yet remain motivated due to a perceived obligation to learn English.

5.6.3 How could integrating such a tool into a CALL system help to enable personalised practice?

An important goal in this study was to gain insight into how integrating a reflective activity such as the one developed might be used to support motivation-based personalised practice into a tutorial CALL system. Now that we have analysed the data it has produced, we can discuss what we have learned and how such an activity could be used.

An advantage of integrating the motivational activity in the present CALL system is that it could provide us with a means of investigating the links between the ideal L2 self types and learners' experiences with the system. For example, in evaluations of learners affective states while completing oral grammar practice with the system, we found that learners tended to enjoy the exercise and be willing to practice with it (Chapter 3, Chapter 4). A review of the profiles with attention to the relative importance of speaking an L2 provides a possible explanation. We see that Type 1+ and Type 3 are characterised by high desiredness ratings for a future in which one speaks Dutch fluently. We also see that Type 2+ is characterised by a moderate desiredness rating for speaking Dutch fluently, in combination with a high rating for speaking foreign languages. Further, we see that all three types are characterised by high desiredness for a future in which one feels comfortable while speaking Dutch, and moderate to high desiredness for a future in which one is confident enough to begin conversations in Dutch. Lastly, we observed that all three types assign higher desiredness to becoming able to speak Dutch fluently versus becoming able to write Dutch fluently. Thus, one plausible explanation for learners' positive experience with the oral L2 practice with our system is that speaking an L2 has a prominent role in learners' ideal L2 selves.

A potential benefit of including such an application in a tutorial CALL system is that we obtain information related to the diversity of the motivational characteristics of our user base. The information on ideal L2 self types presented above is relevant for personalising practice because it provides evidence suggesting that the learners of Dutch we encountered in our experiments were a motivationally diverse group and that, should technology providing motivationally-adapted practice activities become available, this would likely benefit the users.

Besides informing us of the diversity of our learners, the application could also be used to support different strategies for improving motivation. A natural application of the data obtained from the activity would be to identify areas the learner is interested in and adapt practice so that it is perceived as being more relevant to the learner's ideal L2 self-guide. Of course, the case for this type of strategy would be made stronger if there was evidence linking motivational types with different affective experiences during practice or different learning outcomes. One way this could be investigated would be to employ the types as an independent variable and compare affective experiences. For example, the majority of the videos and question-and-answer activities used in our system

depict daily Dutch life, culture and interactions. Therefore, we might expect learners associated with Type 2+ and Type 3 ideal selves, which show more cultural interest and integrativeness, to have a tendency to enjoy practice more.

As a second strategy for enhancing learning experiences, we suggest using learners' motivational profiles to identify areas where they are lacking motivation. Both Csizér and Dörnyei (2005a) and Papi and Teimouri (2014) have shown that learners with a more well-rounded future outlook valuing both social and cultural aspects of the future as well as potential pragmatic benefits report higher levels of intended effort and motivated behaviour. Thus, one motivational strategy that becomes possible by integrating a motivational activity such as the one developed is identifying areas of motivation that are weak and targeting those areas with interventions.

A tutorial CALL component could be built that employs the above two strategies. In the first strategy, learners' areas where they are positive could be used as input to adapt practice so that it is perceived as relevant. This could be expected to enhance learners' motivation to practice. In the second strategy, the system could work at improving the areas of motivation where a learner is lacking. Progress in this area could help learners to develop a more well-rounded ideal self, which has been linked to more intended effort. While it may be difficult for such a program to modify pre-existing beliefs or attitudes the learner brings to CALL practice, it could also be that the learner has not considered the possibilities before. A system could provide exercises designed to let the learner consider a variety of possible futures with the aim of finding elements that appeal to the learner and enhance their motivation in areas that are lacking.

A problem with the strategy described above has to do with the feasibility of developing personalised content. Analysing the prevalence of the ideal self types provides a way to evaluate the feasibility of the work involved in developing additional motivationally-relevant content. Our analysis here suggests the presence of three major perspectives, each covering approximately 25 % of the participants, with 10 individual participants having ideal selves that were not associated with the major types. While this is still far from the worst-case scenario of one perspective per individual, it is clear that developing motivationally-relevant content for the three major and ten additional ideal L2 self types is a substantial amount of work, even acknowledging that some of the content might be shared between the types. In the classroom context, this amount of additional work is unlikely to be feasible. This suggests that technology facilitating the creating of personalised content (Walkington & Bernacki, 2014), by for example, accepting motivational configurations as input and identifying relevant reading material (e.g. Heilman et al., 2010), could help to make individualised practice more feasible.

5.7 Conclusions and future directions

The present study explored the possibility of developing a reflective web activity for eliciting information about learners' L2-related self-guides. We developed a new web-based instrument implementing an innovative card sorting user interface. As an example use case, we employed this information to look at the contents of adult Dutch learners' ideal L2 self-guides using a Q analysis. Three major types emerged from the analysis, each appearing to aspire to speak an L2 in the future, but differing in the importance they assign to becoming competent and functional in Dutch, accessing new social opportunities, and becoming an advanced Dutch user. We observed that the three types were equally prevalent among our learners, and that learners not associated with these three types tended to have ideal L2 selves with a distinct configuration. Due to the small sample size and focus on young adults and professionals in the present study, it is not possible to say whether these findings generalise to the larger population of Dutch learners. At the same time, we note that our participants were a diverse set of individuals, with a wide variety of linguistic and cultural backgrounds, while previous studies on motivational types have tended to focus on a relatively linguistic and culturally homogeneous group of learners (e.g. Csizér & Dörnyei, 2005a; Irie & Ryan, 2015; Papi & Teimouri, 2014). In any case, it seems plausible to conclude from these findings that the contents of learners' ideal L2 self-guides are likely to be similar in some aspects, but different in others. Future work with a larger number of participants could attempt to investigate which facets of an ideal L2 self-guide tend to be shared, and which facets tend to vary with the individual.

An important outcome of the present study has to do with individualising L2 practice. In the L2 classroom, providing individualised practice can be challenging because of the extra workload this places on teachers. Furthermore, classroom time may be more effectively used for group or whole class activities that promote authentic communication. Such learning situations may not be appropriate for individualised learning. On the other hand, tutorial CALL environments have been developed that provide self-study L2 practice individualised to aspects of the learner's developing linguistic knowledge (e.g. Heift, 2016), by logging and analysing past performance in student models. In the literature, we also see that some systems have implemented student models that combine information on linguistic knowledge and performance with information on learners' affective characteristics (Heilman et al., 2010). The implementation of the reflective web activity currently stores information on learners' ideal L2 selves in a student model, making this information available for a future system module that adapts practice. In the long-term, we plan to use this activity to explore to what extent prompting learners to think about their ideal L2 selves can enhance their motivation and contribute information useful for personalising practice.

5.8 Appendix

5.8.1 Participants

Analysis of the background questionnaire data revealed that the group of 59 participants was culturally and linguistically diverse. There were 23 men and 36 women. Their ages ranged from 19 to 62 ($M = 28.5$, $SD = 8.7$). The large majority had post-secondary education (50), another six indicated having between four and six years of public school, one indicated they had pre-university education, another reported elementary school as their highest level, and there was one missing value.

The land of origin as well as first languages were diverse. Group members came from Spain (8), Germany (7), Romania (5), Italy (4), Iraq (3), Russia (3), Poland (2), Iran (2), The Philippines (2), China (2), Switzerland (1), Turkey (1), The Czech Republic (1), Singapore (1), Rwanda (1), Uganda (1), Nigeria (1), Ghana (1), Lithuania (1), The Republic of Latvia (1), Japan (1), Hungary (1), France (1), England (1), The Dominican Republic (1), Canada (1), Brazil (1), Belgium (1), China (1), and Belarus (1), and one unknown.

In addition to their mother tongues, they also spoke a variety of second or foreign languages (participants were told to provide all the second or foreign languages they spoke), with English being the most common (55, along with 4 native speakers), followed by French (13) and German (12). Other foreign or second languages spoken included Spanish (5), Italian (3), Russian (2), Latin (2), Yoruba (1), Swahili (1), Portuguese (1), Latvian (1), Ishan (1), Indonesian (1), Euskera (1), Danish (1), Chinese (1), Catalan (1), and Arabic (1), with two participants indicating that they did not speak a third language and one value missing.

With respect to their reasons for coming to the Netherlands (they could indicate multiple reasons), most learners reported coming to study (24) or work (21) and to accompany a partner (18). Less common reasons were to seek asylum (3), as part of a ‘stagejaar’ (1), for medical reasons (1), because “I find myself here”, while one participant opted not to share this information.

Participants also provided an indication of how much time they were practising Dutch on a daily basis. At the time of the experiment, they reported speaking and listening between one and four hours per day ($M = 2.0$, $SD = 0.85$), with the majority spending two or less hours (67.8 %). The data for reading and writing practice is similar, with participants indicating they spent between one and four hours per day ($M = 1.81$, $SD = 1.0$), and the majority spending two or less hours (64.4 %).

5.8.2 Loading of participants on to factors

Table 5.4: Loading of participants onto their factors, as produced by the inverted factor analysis procedure. Two criteria had to be met to be considered associated with a factor: C1) the loading is significantly high (> 0.37 , $p < .05$), and C2) the square of the factor loading is larger than the sum of the squared factor loadings for all other factors. The threshold for loading significantly is calculated by $1.96 * (1 / \sqrt{N})$, where N is the number of cards, producing 0.37 in the present study. In total, 12 confounded card sorts were observed.

Participant	PC1	PC2	PC3	Loading Result
4056959	0.75	0.22	0.09	Type 1+
3705933	0.73	0.07	0.29	Type 1+
4055722	-0.72	0.28	-0.17	Type 1-
4064361	0.71	-0.05	0.29	Type 1+
4055777	0.66	0.21	0.14	Type 1+
3522101	0.66	0.23	0.17	Type 1+
3355531	0.65	0.16	-0.49	Type 1+
4055507	0.64	0.35	0.26	Type 1+
4055509	0.62	0.1	0.55	Type 1+
2521982	-0.61	0.05	-0.03	Type 1-
2317908	0.54	0.23	0.3	Type 1+
2283203	0.52	-0.13	0.42	Type 1+
4067605	0.52	-0.14	0.25	Type 1+
3341561	0.5	0.15	0.1	Type 1+
2317907	0.49	-0.1	0.27	Type 1+
2283499	0.49	-0.43	0.16	Type 1+
2471765	0.46	0.46	-0.02	Type 1+
3705609	0.44	0.23	0.23	Type 1+
3374167	0.07	0.89	0.03	Type 2+
2473754	-0.11	0.79	-0.05	Type 2+
2379498	-0.31	0.78	0.23	Type 2+
4062739	0.04	0.77	0.06	Type 2+
2376493	0.26	0.75	-0.02	Type 2+
2320857	0.22	0.68	0.36	Type 2+
4062919	0.36	0.67	0.15	Type 2+
4067931	-0.19	0.67	0.3	Type 2+
4067607	0.13	0.65	-0.17	Type 2+
4142811	0.42	0.59	0.17	Type 2+
2320855	-0.54	0.56	0.08	Type 2+
4096359	0.33	0.48	0.27	Type 2+
3355532	0.03	-0.4	0.15	Type 2-
2617006	0.37	0.39	-0.06	Type 2+
2611236	0.13	0.38	0.16	Type 2+

3705930	-0.06	0.37	0.24	Type 2+
3394930	0.14	0.01	0.79	Type 3
2317892	0.43	0	0.7	Type 3
2283202	0.14	0.05	0.69	Type 3
4055506	0.5	-0.21	0.69	Type 3
4062740	-0.02	0.23	0.68	Type 3
2283275	0.11	0.41	0.63	Type 3
2611237	0.27	0.12	0.63	Type 3
2326120	0.17	0.02	0.55	Type 3
2368586	0.51	0.01	0.54	Type 3
4055508	0.25	0.42	0.53	Type 3
2283363	0.37	0.15	0.53	Type 3
4064360	0.1	-0.01	0.51	Type 3
2317855	-0.16	0.28	0.49	Type 3
3355937	0.33	0.22	-0.03	Confound - C1
4067606	0.28	0.04	-0.11	Confound - C1
2376921	0.26	0.21	0.23	Confound - C1
3341562	-0.25	-0.15	-0.06	Confound - C1
3705932	0.28	0.47	0.38	Confound - C2
3374168	0.07	0.34	0.14	Confound - C1
4055510	0.07	0.33	-0.04	Confound - C1
4142812	-0.06	-0.08	-0.03	Confound - C1
2388110	0.41	0.3	0.49	Confound - C2
2376351	0.35	0.29	-0.42	Confound - C2
3355533	0.33	-0.03	0.33	Confound - C1
2617004	-0.14	0.29	0.29	Confound - C1 and C2

5.8.3 Card sorts for the different ideal L2 self types

Table 5.5: The items and their scores for the types (factors) that emerged from the analysis. Scores for a type were calculated from participants who load on the type significantly. The score for an item is the weighted mean of that item, where the weight is the loading coefficient representing the degree of similarity between a participant and the factor. Scores have been scaled to the range [5, 100] to match how they would have appeared in the instrument during the sorting activity.

Possible Self Item	Type 1	Type 2	Type 3
I have made people in my life happy or proud by learning Dutch (1)	33.84	17.74	14.10
I can communicate with native speakers in Dutch without the help of others (2)	90.98	72.63	58.97
I feel comfortable when I speak Dutch (4)	81.61	69.21	76.81
I have people in my life who think it's great that I study Dutch (6)	41.59	16.48	18.83
I feel confident enough to begin conversations in Dutch (9)	63.69	69.55	87.49
I can write fluently in Dutch (10)	72.01	22.59	82.27
I speak Dutch fluently (11)	78.76	49.95	100
I feel comfortable speaking Dutch in a group (of 5 or more people) (14)	77.07	58.64	56.66
I can express my emotions in Dutch (e.g. happiness, anger, sad, love) (16)	78.28	58.31	66.8
I can understand Dutch slang and use it properly (17)	53.62	30.78	63.35
I can express myself creatively in Dutch (20)	60.27	52.85	83.28
I can function in my day-to-day life in the Netherlands without ... (21)	85.64	72.16	5.00
I can use Dutch to get a better job (22)	76.92	44.91	59.9
I speak a foreign language often (24)	20.1	85.58	31.93
I know a number of different foreign languages (25)	34.00	88.00	28.59
I can use Dutch to get a better education (26)	59.49	34.7	60.45
I often meet people from a variety of different countries and cultures (30)	10.88	81.62	40.22
I have opportunities to learn about the values and customs of (35)	19.41	71.81	33.89
I get along well with Dutch people (36)	37.18	74.69	64.18

I have opportunities to meet Dutch people (37)	26.09	47.96	46.56
I spend a lot of my time with Dutch people (38)	32.79	36.41	32.50
I have opportunities to spend time with other Dutch learners (39)	14.00	25.56	32.46
I often watch Dutch films or tv programmes (42)	33.56	19.1	31.77
I often read Dutch books, magazines, or newspapers (43)	26.71	20.71	27.09
I am similar to Dutch people (44)	16.14	7.14	12.99
I have good friends who are Dutch (46)	32.21	59.01	56.6
I live in the Netherlands permanently (47)	35.44	6.34	27.17
I understand Dutch people and their way of life (48)	49.37	47.24	41.78

CHAPTER 6

Discussion and conclusions

There is no doubt in our minds that future research should be moving toward increasing integration between: (a) group-based quantitative approaches representing a macro-perspective and (b) individual-centered social approaches representing a situated, micro-perspective. (Ushioda & Dörnyei, 2012, p.405)

L2 motivation research has changed over the years. It began with Gardner's work on the relationships between attitudes and motivation, taking a very macro perspective not attached to any particular context. In the 1990s, Crookes and Schmidt (1991) argued for more educator-friendly perspectives, delineating different levels of context such as the classroom and the micro-context. In recent years, micro-perspectives have become associated with sociocultural perspectives which highlight learner-context interactions as a process through which L2 motivation emerges (Ushioda, 2009b). Interaction in sociocultural theory is typically thought of as between two humans. However, learners can and do also learn languages in contexts which do not involve learner-other social interactions. The homework that is part of a typical language course, while often completed in groups, is just as often completed alone; listening activities are often completed alone. Self-study CALL systems are a third practice context where learning can take place without human-human interaction. As these types of environments become more popular, and judging by the rate of adoption of technology this seems likely, there will also be a need for motivational perspectives that are individual-centred and sensitive to learner-context interactions, but do not necessarily have a strong social component.

This thesis is concerned with L2 learning that takes place in a computerised learning environment. Its overall aim is to contribute to the understanding of L2 learning processes in such an environment. This goal is approached with research focused on affective dimensions such as emotion and motivation and their interaction with language learning.

More specifically, the research in this thesis pursued a number of goals. First, it aimed to take stock of developments in recent L2 motivation research. In Chapter 2, we surveyed practices from CALL research focusing on affective dimensions of learning and compared current practices in the CALL literature with the ideas from recent developments in L2 motivation literature, with the goal of evaluating the degree of match and articulating directions for future research. Second, it attempted to instantiate the abstract ideas put forth in recent perspectives in L2 motivation research (Ushioda, 2009b) and address the directions for future research that were proposed in Chapter 2 by developing technology to operationalise said ideas in the context of a CALL system that would provide practice and feedback to improve oral proficiency. Third, it employed this technology to conduct research on L2 learning conditions. This was realised as research into affective states during practice in the promising, but little-tested context of a CALL system providing oral L2 self-study and employing ASR to provide automatic CF. Fourth, our research employed the technology to explore learners' motivational profiles. This was realised as research into the diversity of learners' ideal Dutch L2 selves.

A key challenge for the present research was how to go about operationalising ideas about situated perspectives to motivation. Ushioda herself uses a qualitative methodology, while our perspective was on how to develop technology that would allow these issues to be explored in a computerised self-study environment that was in itself challenging to develop. In other words, our concern was how best to exploit a computer environment's potential to realise research conditions for situated approaches to studying affective dimensions of L2 learning with a focus on speaking practice.

The remainder of this chapter is structured as follows. In section 6.1 we reflect on our approach to implementing a situated look at affective dimensions in L2 learning in a computerised environment. Section 6.2 summarises our findings concerning in-practice affect and the impact of CF, practice time, proficiency and learning gains. In section 6.3, we discuss the benefits of employing logs of learner behaviour and the relationships between behaviour and affect. Section 6.4 highlights our results from our exploration of the diversity of Dutch L2 learners' ideal L2 selves. Considerations of the limitations of the present research and future directions are presented in section 6.7, followed by final words in section 6.8.

6.1 Significance of the experimental platform developed

The system we developed for this research constitutes an innovative environment for L2 learning and research. Several reasons have already been presented by Penning de Vries (2015). For completeness, we include some of the points mentioned in that work, but add complementary observations from the present research.

First and foremost, the system we have developed makes speaking practice with automatic CF available to learners as a self-study exercise. This required the identification and implementation of a pedagogical context suitable for ASR. The result can be considered innovative, as the system supports oral word order exercises using ASR to provide automatic CF, a previously unavailable spoken modality practice feature, going some way to providing learners with features typically only available from human tutors. Further, the web-based delivery means it could be accessed by learners anywhere there is access to a computer and a suitable microphone. Second, the system serves as a comprehensive platform for CALL-based experiments. It has been successfully deployed as research platform in studies investigating CF, spoken L2 proficiency development and affect. For researchers, it possesses a number of useful features. First, it logs user and system actions, including all users' spoken productions, allowing for a detailed analysis of learner behaviour and performance. Second, it implements spoken production and receptive Dutch L2 proficiency tests targeting word order accuracy, allowing for research into the effectiveness of the system in developing that aspect of spoken L2 proficiency.

In developing the platform, a software architecture was chosen that allows the definition of task lists composed of L2 task objects. This choice had several benefits. First, task lists allowed researchers to define experimental procedures that are automatically carried out by the system. For example, task lists were used to control proficiency tests by creating different lists in which the proficiency test versions were alternated. Learners could then be easily assigned to an experimental procedure. Further, during an experiment the system managed these procedures. This was accomplished by presenting the learner with informational screens before and after an activity, and which prompted the learner to click on a link when they were ready to begin an activity. This automated the experimental procedure, contributed to consistency in the experiments and reduced likelihood of experimenter mistakes or influence.

An additional benefit of explicitly modelling tasks in the system is that it made the system extensible. By including tasks, we were able to expand experimental protocols tested during piloting to include additional components, such as microphone tests, tutorials, and proficiency test trial exercises. Organising experimental procedures using task objects as a unit also allowed us to integrate the two instruments we developed for eliciting affective data.

Most relevant to this thesis is that the platform allowed us to explore

learners' affective states in oral L2 practice. This required becoming familiar with the various theoretical frameworks and constructs used in psychological research on language learning and presented a fundamental methodological challenge, due to the nature of the phenomena we wanted to study. As Ellis (1994) notes, "It is, of course, not easy to study the role that affective states play in L2 learning in a systematic manner, if only because they are so individualistic and changeable" (p. 479). Developing an approach to studying affective dimensions from a situated perspective or process-oriented view in a CALL system was challenging due to a lack of standard methodologies. Previous research has primarily employed qualitative methods. Although Dörnyei and Ottó (1998) proposed their process model of L2 motivation, to our knowledge few CALL studies have attempted to investigate affective dimensions as they change over time. In answer to this challenge, we developed an in-practice questionnaire employing items referencing the current practice environment (Dela Rosa & Eskenazi, 2011) and targeting aspects of learners' affective state at the time of practice. We defined an experimental procedure that took repeated measurements, made possible by the task list objects described above, to investigate how affect changed with time. Logging features mentioned above allowed us to investigate how affective state related to practice behaviours. At the time of development, we were unaware of any system implementing such situated approach to investigating affect during L2 speaking practice in a self-study computer context.

Equally relevant is the development of the activity for eliciting ideal L2 selves. The activity is a departure from previous quantitative instruments that employ different types of survey items (Hessel, 2015; MacIntyre, Mackinnon, & Clement, 2009) which, while effective for researchers, are typically not designed with the learner experience in mind. Instead, we drew on descriptions of techniques used in counselling (Shepard & Marshall, 1999) and self-directed learning (Pemberton & Cook, 2012) to develop a computer-based activity that combines a set of guided, interactive, tutorial-style instructions with activities that ask learners to browse, select and sort virtual possible self cards. These types of card sorting activities have been reported to be interesting, enjoyable and to promote self-reflection (Pemberton & Cook, 2012). In addition to having potential benefits for learners, the activity also has potential to be useful to researchers. As demonstrated in Chapter 5 and elsewhere (e.g. Irie & Ryan, 2015), card sorting activities can be analysed using Q methodology to identify and compare distinct perspectives on subjective domains such as learner beliefs, perceptions or desirable future selves. The instrument is equally capable of supporting other types of L2 motivation research, such as investigating how different possible self attributes, such as desirability, plausibility and discrepancy with the present self, relate to a possible self's capacity to motivate (e.g. Hessel, 2015; MacIntyre, Mackinnon, & Clement, 2009).

Developing these aspects of the system was challenging due to the high degree of multidisciplinary required for developing such components and the

limited time frame available in a doctoral program. Despite the challenges, we find that we have successfully deployed the platform in our research and suggest that it constitutes a significant contribution to CALL research. First, it provides a way to study affect, which is considered difficult to study, in a context which is under-researched, namely L2 speaking practice with automatic feedback. Second, it contributes a novel method for eliciting L2 selves data. Together, the platform provides an authentic environment for investigating L2 learning effectiveness and processes. It illustrates that computer technology can be used to pursue a comprehensive understanding of L2 learning as it occurs in specific contexts with specific features. An additional advantage of such a system is that it produces a rich corpus of linguistic and meta-linguistic data which can be made available to researchers to study L2 learning from different perspectives. Furthermore, the speech data collected and the corresponding data resulting from ASR processing (segmentations, confidence scores etc.) and from the loggings can be employed to further develop and refine future ASR-based CALL systems. In sum, we find that it is a comprehensive system with strong potential for future research and development.

In the following sections, we reflect on our work using the platform to pursue our research aims.

6.2 What factors influence in-practice affect in ASR-CALL self-study?

The impact of CF

In Chapter 3, we found that learners reported positive affective experiences on the dimensions measured by the in-practice questionnaire. In both the NOCF and CF groups, learners reported enjoying practice with the system, being generally willing to practice and moderate in their expectations of doing well in the exercises. Statistical analysis of how these three affective dimensions changed over time suggested, first, that NOCF and CF groups had equivalent trajectories during a practice session and, second, that differences between the first and last sessions (or lack thereof) were shared between the groups.

Chapter 4 also explored the impact of CF on learner affect, with an analysis that stepped back from the micro-perspective taken in Chapter 3. To obtain a more general evaluation and to increase the power of our statistical analyses, the analysis in Chapter 4 was performed using a pooled data set reflecting the views of the participants who used the system in three different experiments. We reported on a comprehensive assessment of the system by analysing the impact of CF on changes in the target linguistic feature (Dutch V2 word order accuracy), in-practice affect and post-practice evaluations. Interestingly, we found that CF, on its own, had a limited impact on these measures (we detected two interaction effects involving CF which are discussed below). In the case of proficiency, both the NOCF and CF groups significantly improved

V2 accuracy from pre-test to post-test. In-practice measures of affect were also found to be equivalent. On the general evaluations as reported on a post-practice questionnaire, we found the groups differed in the amount of difficulty they experienced. First, the NOCF group found practice to be more difficult than the CF group. A more detailed comparison of the NOCF, CF-P (prompt feedback) and CF-R (explicit recast feedback) groups revealed a difference between the NOCF group and CF-R group, which received feedback that revealed the correct form. Summarising these findings, together chapters 3 and 4 contribute evidence showing that CF did not have a negative impact on learners' affective states in the current practice context.

What contributions do these findings make to our understanding of CF? As discussed in Chapter 3, to date, most recent research into CF has investigated the effects of CF on developing L2 proficiency, perhaps in answer to some strong views asserting that CF is ineffective in facilitating L2 acquisition (Krashen, 1982; Truscott, 1996, 1999). There is now increasing evidence that CF has a facilitative effect, and calls have been made for research to begin to untangle the various factors mediating CF effectiveness, with one important variable likely to be the learning context (Lyster et al., 2013). A second, often-mentioned variable implicitly assumed to influence CF effectiveness is the manner in which learners respond affectively to the feedback they receive: teachers and researchers are often cautious about the potentially negative impact of CF on learner affect (e.g. Ellis, 2009; Raimes, 1991; Vásquez & Harvey, 2010). In a recent special issue on CF, Ellis (2010) noted that surprisingly few studies had addressed the topic.

At present, CALL research investigating the affective or motivational impact of CF in self-study environments is an innovative area, with relatively few studies. One exception, discussed in Chapters 3 and 4 but worth returning to here, is the work by Cornillie, Clarebout, and Desmet (2012), which has interesting similarities with the work completed in this thesis. Both studies have in common that they investigated the impact of CF on learners' affective experiences in CALL practice. Furthermore, although they used different instruments, there appear to be similarities between the affective constructs measured: their game experience would seem to compare to our enjoyment, possibly together with willingness to practice, on the basis that learners who enjoy and are willing to practice would seem to be having positive experiences. Similarly, the study's perceived competence appears comparable to our measure of self-efficacy. It is interesting then to note that Cornillie et al. (2012) find that immediate CF can be provided without negative impact on learners' game experience, even in the case of relatively low performing learners who receive CF frequently. In Chapters 3 and 4, we also found that CF can be provided frequently without negative impact on learners' affective practice experience. Furthermore, our behaviour logs analysis in Chapter 3 shows that learners tended to improve in performance, which corresponds with Cornillie et al.'s suggestion that learners' positive perceptions of CF were perhaps attributable to their improvement during practice. Interestingly, our findings in Chapter 4

also provide support for the other possibility mentioned in the study, i.e. that learners “perceived competence was unaffected by CF” (p. 273). We will return to this below when discussing the impact of proficiency and learning gains on learners’ affective experience.

In our review of the broader language learning literature, we found that previous research has explored CF and affect by either surveying learners’ beliefs and perceptions regarding CF while assuming the classroom as a context (e.g. Agudo, 2013; Kartchava & Ammar, 2013; Loewen et al., 2009), or employing a situated approach to investigate the CF-affect relationship in the classroom (e.g. Busse & Walter, 2013; MacIntyre et al., 2011). In both cases, the results have been somewhat mixed, with learners tending to believe CF to be effective yet acknowledging that receiving CF can sometimes be unpleasant. At the time of conducting the research, to our knowledge, only MacIntyre et al. (2011) had conducted research related to CF and affect in an oral practice context.

The findings from Chapter 3 and Chapter 4 thus contribute to the CF literature by documenting the impact of CF in a novel oral L2 practice context, an ASR-enabled CALL system for spoken grammar practice, from an affective point of view, and identifying a practice context where CF does not appear to have a negative impact.

Although learners tend to find speaking an L2 to be anxiety provoking, the practice context in the present study was significantly different from contexts previously researched in the literature where speaking and anxiety have been investigated. The practice context in our study was deliberately designed to provide individualised CF in such a way that anxiety would be minimized. Prior work has indicated that learners feel comfortable in computer-based L2 practice (Warschauer, 1996). In the computer-based self-study offered by our platform, learners speak with the computer and not with another individual. This implies that there is no risk of losing face due to errors, the learner is free from social timing constraints and can take time when formulating an utterance. While at present ASR-enabled CALL self-study contexts such as the one we investigated are relatively rare, the availability of such systems is likely to increase in the future as researchers document their pedagogical effectiveness and develop methods for making their development more feasible (e.g. Presson et al., 2013).

A second point to consider is that surveys in the literature have suggested that learners have positive attitudes towards CF and find that they do not have sufficient opportunities to receive it. This would seem to suggest that the learners in our study who receive immediate CF on word order errors would report more positive experiences, while we did not detect this. There are several plausible explanations for this outcome.

First, it has been documented in the literature that learners’ general preferences and perceptions obtained with surveys do not always match with their actual preferences in particular learning situations. Plonsky and Mills (2006) found that, prior to beginning specialised training related to CF in

the classroom, learners expressed a preference for having all of their errors corrected, while at the end of the training, after reviewing examples from their classroom, learners' views shifted to recognise that CF can have a negative impact and that not all errors should be corrected. Similar findings were reported by Cathcart and Olsen (1976). It is possible that, in the experiments reported on in this thesis, learners experienced something similar to the students in the previous classroom studies. While CF may appear to be an attractive feature to learners and while they may express a *general* preference for CF, for example, before or after practice has completed, it may be the case that learners react more ambivalently *during* practice with feedback, which reported every word order error made.

A second plausible explanation is that this outcome may be related to the amount of time learners practised with the system. As Jamieson and Chapelle (1988) points out "As students enter the learning situation, those with good attitudes toward L2 acquisition tend to practice using any means that are readily available, useful, or interesting (Marty, 1981)" (p. 157). In the experiments, learners practised for a total of 90 minutes over two weeks, during which they had opportunities to receive L2 input in the form of video clips and questions, and opportunities to produce output, differing only in the presence or absence of CF. It is possible that since both experimental conditions offered a rich learning context, it was difficult for participants to appreciate the added value of CF in such a limited lapse of time.

A third plausible explanation is related to learners' prior experiences using computer technology for language learning and their expectations for practice during the experiment. It is reasonable to assume that participants in our experiments had experience using computers, as the majority indicated this in their background questionnaires. It is also likely that they had some experience using computer technology for language learning, as the course they were recruited from had an online learning system component and had a computer-based testing component. However, the current state of CALL environments typically used in language labs is such that very few have the ability to provide automatic CF on learner speech by means of ASR. Further, during recruitment of the participants, we did not reveal that two different versions of the system were being tested. Thus, we find it plausible that participants in our experiments did not expect to receive CF and that therefore the absence of CF may not have had a role in their affective response towards practice with the system.

Above, we hypothesised that the combination of the computerised context and learners' expectations for CALL software influenced the affective experiences observed. While this hypothesis should be tested in future research (see our discussion of limitations below), in the present research we are able to draw conclusions regarding two factors which influenced affective state during practice.

Time-on-task

The methodology employed in this thesis provides a rare dynamic view of affective dimensions of learning in a CALL environment. In the present research, we found that learners' enjoyment, willingness to practice and self-efficacy changed over the course of the experiment. In Chapter 3, we examined the affective trajectories of learners within a 45-minute practice session. The analysis showed different changes for different constructs. For enjoyment, the initial sample, which was taken after a tutorial introducing the system, but before practice began, was lower than samples taken later during practice. These results suggest that learners had a positive experience during practice which exceeded their initial expectations. For willingness-to-practice, the pattern was the opposite. We found that learners were significantly less willing to practice at the end of the session. For self-efficacy, we found a modest rising trend, with learners reporting more confidence in their abilities at the end of the experiment relative to when practice began. These findings are among the first to document fluctuating affective state during spoken practice with an ASR-enabled CALL system.

The findings contribute to other research which has shown that WTC, anxiety and approach-avoidance motivation can fluctuate during speaking tasks (e.g. MacIntyre & Legatto, 2011). However, it is worth pointing out that the present research had different methodological characteristics compared with recent research into dynamic affect (e.g. MacIntyre et al., 2011). In those experiments, task difficulty was varied deliberately to trigger and observe changes. In our system, we suggest that practice was more naturalistic. We did not deliberately attempt to trigger changes, but monitored them as they evolved in practice that was intended to be beneficial and to have a positive impact but not designed to elicit particular changes. In the work by MacIntyre et al. (2011), the time scale is much smaller, with learners recording moment-to-moment changes via stimulated recall procedure that occurred after the tasks had been completed. Our approach was to elicit learners' affective state during practice, every 15 minutes, with this interval chosen to minimize any potential negative effect of the research instrument. Another difference is that MacIntyre et al.'s method tracks one affective dimension over time and is intended to analyse dynamic changes for a single individual. Our approach allows the tracking of three different dimensions, which correspond to the tripartite structure proposed by Dörnyei (2009a) consisting of affect (emotions), conation (i.e. motivation) and cognition. The changes over time reported above suggest that the constructs are sensitive enough to be applicable for educational (i.e. learning conditions) research.

The methodology also allowed an assessment of changes in affective dimensions resembling a more conventional pre- and post-test design. In Chapters 3 and 4, we compared affective states on the first and last practice sessions. Chapter 3 reported rather positive findings, indicating that enjoyment was high in the first session and remained stable, and that both willingness-to-practice

and self-efficacy increased from the first to the second session. Chapter 4 conducted a similar analysis on a data set containing a larger number of participants pooled together from Chapter 3 and two other experiments with comparable procedures. In Chapter 4, we found that learners' affective states were similar in the first and second session: the data indicated that learners enjoyed and were willing to practice, and that the difficulty level was appropriate. This is a positive outcome, as it is also possible that practice in CALL environments can lead to negative experiences (e.g. Raby, 2007; Van Aacken, 1999). Together, these results suggest that the spoken language practice provided learners with a generally positive learning experience that persisted after repeated practice.

Proficiency and learning gains

Chapter 4 contributes to previous CALL research that has shown that proficiency can differentially influence affective learner characteristics (e.g. Izquierdo et al., 2015; Jauregi et al., 2012). In Chapter 4, one of our aims was to stand back from the detailed approach in Chapter 3 and investigate learner affect as it relates to developments in learners' Dutch V2 accuracy. In the second part of the analysis, we adopted initial V2 accuracy and learning gains as independent variables and compared learner trajectories for enjoyment, willingness-to-practice and self-efficacy.

The analysis revealed that the relationship between V2 learning gains and in-practice affective experiences was mediated by learners' V2 proficiency at the start of the experiment. Medium V2 proficient learners who improved reported positive affective states that were roughly equivalent in the two sessions, while medium V2 proficient learners who did not improve tended to report less enjoyment, willingness-to-practice and self-efficacy in their final session. On the other hand, we found that low proficiency learners reported positive affective states in their first and final sessions, both when they improved and did not improve. Put another way, we found that medium V2 proficient learners had affective experiences that changed in a way that roughly followed the changes in their V2 proficiency, while the low V2 proficient learners' affective experiences did not always resemble the changes in their V2 accuracy.

To explain these findings, we suggested that the affective differences observed between the low and medium proficiency groups could be explained by difference in the amount of resources each group had available to gauge their progress. We speculated that medium proficiency learners had more resources to gauge their progress and became critical if they felt they were not improving. The self-efficacy results showing lower values for the medium proficiency learners who did not improve could be interpreted this way.

Chapter 4 also included an analysis of learners' post-practice evaluations of the system. These results showed that, although CF did not appear to impact in-practice affective states, following practice, low proficient learners who practised with CF tended to find practice more attractive and easier

relative to their counterparts who practised without feedback.

Together, the findings in Chapter 4 show that individual characteristics, properties of the learning environment, and time interact to impact learners' affective learning experience.

Now that we have reviewed the findings from Chapter 4, we can return to the study by Cornillie et al. (2012) to discuss a second possible interpretation of that study's finding concerning the apparent lack of impact of CF on learners' perceived competence. Recall that participants in the study formed positive perceptions of CF despite their low performance in the game and likely having received large amounts of feedback. To explain this, the authors present two possibilities: "either students' perceived competence was unaffected by CF, or they improved as a result of such feedback" (p. 273). The findings from Chapter 4 appear to connect with this point nicely, and seem to suggest that the apparent lack of connection between feedback and perceived competence might be due to learners' proficiency. If, in the study by Cornillie et al. (2012), the low level of game performance can be taken as an indication of learners' L2 proficiency level, then it might be possible that their performance did not improve but that they nevertheless had a positive game experience, as was found with the low proficiency learners in Chapter 4. In fact, if we consider the findings in Chapter 3, it would seem both explanations are possible. In our view, an interesting direction for future research would be to investigate what criteria learners attend to when judging their competence in CALL-based practice with CF, and how these criteria vary with different proficiency levels.

6.3 The contribution of behaviour logs to understanding affective states

With the behaviour logs used in Chapter 3, we were able to see that CF had an impact on learner behaviour and monitor how behaviour changed with time. In practice without CF, learners tended to make one attempt at producing a spoken answer to the questions they received. Further, this behaviour pattern remained stable throughout the experiment. The effect of adding CF that required learners to correct their errors before moving to the next question was that learners produced multiple attempts per question (i.e. exerted more effort). Initially, they produced an average of two attempts per question. However, over time, the CF group was able to complete questions using increasingly less effort. In our analysis of the log data we also considered in-practice performance, which we operationalised as the number of questions answered correctly on first attempt. Here, the logs allowed us to see that performance of both groups was equivalent, increasing from session 1 to session 2 for both.

Having this information about logged behaviour complements the affective in-practice data. It allows us to observe that, although the trajectories of learners' affective states along the dimensions we measured did not differ, the learner behaviour during practice did differ, showing that the CF group

received many instances of feedback during practice and as a result produced more spoken utterances. We find that CF stimulated learners to produce more spoken utterances without impacting affective dimensions of the learning process. However, concerning in-practice performance, CF did not provide a significant advantage relative to the NOCF group, as both groups were able to improve. A second benefit of the log data was that it allowed us to investigate the relationship between learner affect and behaviour. We found a significant positive relationship between performance and self-efficacy in the CF condition only, suggesting that learners may use CF to monitor their performance and inform their self-efficacy beliefs.

According to Felix (2008), investigating students' perceptions of their learning environments is one of the most researched areas in CALL. However, some research has shown that perceptions can be inconsistent with objective measures of practice behaviour (e.g. Fischer, 2007). Further, Monterrat, Desmarais, Lavoué, and George (2015) found that when human experts adapt a learning condition to either match or deliberately conflict with learners' preferences, this can significantly impact learners' time spent using the system, yet not be mentioned in learners' self-reports of their affective experience during practice.

The finding that self-efficacy and performance were significantly related suggests that CF can influence how learners evaluate their performance. Feedback provided in the system was explicit and, given the design of the interaction which required learners to dismiss feedback popup windows provided by the system, is likely to have drawn learner attention. As learners in the CF condition improved their performance, they would have seen the negative feedback less frequently and the confirmatory green check-mark more frequently. This description of the process seems to be a plausible explanation of the link between self-efficacy and performance in the CF condition. It would seem that CF had a role in regulating learners self-efficacy beliefs. Learners practising in the NOCF condition obviously did not have access to CF. In the absence of external CF, they were left on their own to appraise their ability to do well in the exercises. The lack of an external 'standard' reference means that learners were free to decide on their own criteria for evaluating their ability to do well. At the same time, a number of studies have shown that learners can have difficulty detecting the errors they make (indeed, pushing learners to notice their errors is argued to be one of the functions of feedback, e.g. de Bot, 1996). Further, as the pedagogical goal of practice was never revealed to learners, and some learners indicated that they felt they had improved in pronunciation and vocabulary, it is also plausible that learners varied in how they judged their progress, and that the criteria they used at least sometimes differed from word order. These possibilities, learners' inaccuracy in judging their errors and variation in the criteria they used to judge their progress, are plausible explanations of lack of a strong relationship with word order accuracy in the NOCF condition.

A separate contribution of the practice logs is that they provide a possible

explanation for the low willingness-to-practice levels observed at the end of a practice session. One plausible explanation for this is simply that learners are tired, having completed L2 Dutch practice for 45 minutes. However, there is a second explanation, which, without the availability of behaviour logs, would not have been possible to suggest. In reviewing the logs, we observed that during the last 15 minutes of practice, learners frequently completed questions they had already visited before. This was due to the design of our practice which controlled for time-on-task by returning learners who moved more quickly through the material to the start of the exercises and asking them to continue practice. Based on our data, it is difficult to disentangle the effects of time-on-task and repetition. We find it plausible that learners would lose interest in repeating the material and become less willing to practice. However, we also think it is plausible that learners may appreciate a second chance to go through the material. This could help with listening comprehension during the videos, or additional attempts at the questions (especially in the CF condition where learners were automatically forwarded to the next question after making a correct attempt).

6.4 Profiles and prevalence of Dutch learners' ideal L2 selves

The work in Chapter 5 began with the assumption that learners' may have different long-term motivational characteristics (Csizér & Dörnyei, 2005a; Papi & Teimouri, 2014), and that these characteristics can play a role in learners' affective experience or behaviour in L2 practice (Dörnyei, 2002; Julkunen, 2001). The general idea underlying our work is that learners with different ideal and ought-to L2 selves may have different affective responses to the spoken practice offered by our platform.

A prerequisite step in investigating the links between the ideal and ought-to selves and affective experience with the system was obtaining a method capable of characterising these selves. At the time of beginning the work, only the instrument by MacIntyre, Mackinnon, and Clement (2009) had been designed to probe different facets of a learner's ideal L2 self (Hessel, 2015). In considering possible approaches to developing an instrument in our platform, it seemed that implementing a standard questionnaire component would miss an opportunity to exploit modern technology to provide learners with a more interesting experience when reporting their data. A review of methods in broader literature working with possible selves revealed a procedure called the Possible Selves Mapping Interview (Shepard & Marshall, 1999) which made use of card-related activities and which seemed as if it would be both more engaging to learners (for some evidence of this, see Pemberton & Cook, 2012) and exploit the capabilities of modern web technologies.

Chapter 5 presented the resulting novel computer application we developed, showing how it can be used to study the motivational characteristics of a group

of adult Dutch L2 learners who participated in the experiments described in Chapters 3 and 4. Adult L2 Dutch L2 learners are a population which, to our knowledge, has not yet been studied in the L2 motivation literature. In the study, learners were asked to use the online card sorting application to consider different possible futures. They browsed, selected and sorted possible self cards and in the process rated different possible selves along the dimensions of desiredness, self-efficacy and plausibility. The desiredness ratings were submitted to a Q analysis, (employing an inverted factor analysis) to investigate the diversity of ideal L2 self types and the feasibility of representing our population of learners with a relatively smaller number of types.

Analysis of the card sort data produced 13 unique ideal L2 self profiles, of which three were analysed in detail because they were much more prevalent among the learners (see below). Inspection of the composite card sort for Type 1+ suggested that learners of this type prioritise becoming functional in Dutch and value the pragmatic benefits associated with Gardner's (1985b) instrumental orientation more than the other types, yet are indifferent towards potential social or cultural future benefits such as having Dutch friends or meeting people with different cultural backgrounds. Type 2+ was found to be drawn to a future in which they are proficient in and have opportunities to use foreign languages, as well as meet people from other countries and cultures. While this type can be characterised as having a strong international posture (Yashima, 2009), at the same time their Dutch-specific interest was weaker than the other types. Type 3 appeared to be most attracted to a future in which they have become proficient and confident Dutch speakers. The group was observed to react moderately to other possibilities such as potential career benefits or social possibilities, often with scores appearing between Type 1+ and Type 2+. We interpreted this to mean that they had a stronger motivational base and that their hopes were more well-rounded. The three types were also in consensus regarding the desiredness of three possible self cards: understanding the Dutch way of life was considered to be moderately desirable; reading Dutch books, magazines and other materials was a low priority; spending time with Dutch people was also a fairly low priority.

Two subsequent analyses addressed the self-efficacy and plausibility ratings of learners' most-desired selves. In a first analysis, we investigated the homogeneity of self-efficacy and plausibility within the groups of learners associated with one of the three types above. This was accomplished by investigating whether the self-efficacy and plausibility ratings could be characterised as coming from two or more distinct populations of learners. We find that these beliefs were homogeneous within the subgroups. In a second analysis, we completed a between-group comparison of the self-efficacy and plausibility ratings (for the most-desired selves). The findings showed that all three groups were similar in their beliefs.

Concerning the prevalence of the types, the findings from Chapter 5 suggest that the population of Dutch learners in our experiments were motivationally diverse in the sense that no one ideal L2 self type dominated

the population. Rather, we found 13 unique types with the three Dutch-related future perspectives described above approximately equally prevalent among the learners (25 - 28%). The remaining perspectives were each associated with one individual only, except in one case where the perspective was associated with two individuals. In the large majority of cases learners were associated with one ideal self type only.

Seen in combination with the findings from Chapters 3 and 4, the ideal L2 self profiles described in Chapter 5 provide an interesting possible explanation for learners' generally positive experiences during oral L2 practice with our system. In reviewing the profiles to search for possible reasons why learners may have had generally positive experiences during oral L2 practice with our system, we noticed that all three types share a strong desire for futures related to speaking an L2: the Type 1+ and Type 3 profiles rate the future in which one speaks Dutch fluently highly. In the Type 2+ profile, we find that the future where one speaks a foreign language is highly desired together with a moderate desire to speak Dutch fluently. The types also had similarly high desirability values for futures related to feeling comfortable while speaking Dutch and feeling confident enough to begin conversations. Lastly, in all three types speaking fluently was rated higher than writing fluently. Thus, we find that learners' positive reactions to the system may have partly been due to learners' aspirations to speak an L2 in the future and the spoken L2 practice offered by the system.

Together, the results from Chapter 5 suggest that the desire to become a proficient L2 Dutch speaker can be found in qualitatively different ideal L2 self configurations. Information about how the desire to learn to speak Dutch fits into the configuration could open doors to different ways to motivating individual students. Furthermore, the findings concerning the prevalence of the different ideal self types have implications for the necessity and feasibility of personalising practice which we return to later, after stating our general conclusions.

6.5 General conclusions

We have developed a comprehensive system that seems capable of supporting research of the type described at the start of this section. We have evidence that in this environment learners find L2 speaking practice enjoyable, are willing to practice, and that their self-efficacy builds over time. Further, immediate, individualised CF on learners spoken utterances can be used to push learners to interact with their language without negatively impacting learner affect. We have some evidence that learner affective experiences are influenced by the combination of prior knowledge of Dutch and the degree to which learners benefited from practice. We also have some evidence that the ideal L2 selves of adult Dutch learners can be grouped into meaningful types, but that learners appear to be motivated to learn Dutch for diverse reasons. Finally,

the development of the card sorting activity provides us with the means to investigate situated affect and behaviour as it relates to learners' ideal L2 self profiles, with the potential for future versions of the platform to employ this information in student model modules that personalise practice.

6.6 Limitations

The research carried out in this thesis had some limitations which are useful to discuss for the purpose of providing avenues or recommendations for future research.

First, we note that the experiments reported on in this thesis were conducted with a relatively small number of participants which placed limitations on the power of our statistical analyses to detect effects as well as the representativeness of the ideal L2 self types described in Chapter 6 with respect to the larger Dutch L2 learner population. Our participants, recruited mainly from a university language institute, were already enrolled in a fairly demanding Dutch L2 course and participating in our experiments required extra effort to schedule appointments and attend practice sessions in a university computer lab. While many students seemed to view this as an extra opportunity to practice their Dutch and participated in our experiments, a similarly large number of students experienced this as an extra burden and opted not to participate. An issue limiting the number of subjects in Chapter 4 is related to the varying levels of proficiency we observed in the target linguistic feature, Dutch V2 word order, in our participants. In Chapter 4, otherwise suitable data was excluded from the analysis to avoid ceiling effects due to learners already being proficient in V2. This was unexpected, as we took deliberate steps to identify proficiency levels likely to benefit from word order practice by consulting with experienced Dutch L2 teachers. Despite these efforts, in our experiments we still encountered relatively large numbers of learners at upper-beginner and lower-intermediate levels (i.e. A2 and B1 CEF) who were already proficient with Dutch V2.

A second limitation of the research is the relatively short duration of students' practice with the system. Hubbard (2005) notes two problems with research that spans only a short duration and its associated novelty effects:

- 1) while novelty can be positive, any positive effects can wear off quickly so that initial favourable results may not be sustained;
- (2) novelty is also associated with confusion and frustration, or just inefficiency, so that initially less favourable results may show improvement over time.(p. 361).

In the experiments reported on in this thesis, students typically completed two practice sessions, each lasting 45 minutes, over the period of two weeks. It is unclear whether this length of time and number of practice sessions was enough to safely rule out novelty effects from influencing the affective states

of the learners. One way to address this issue would be to conduct a more longitudinal study to investigate this possibility.

A third limitation of the study is the lack of qualitative methods employed. Although the in-practice affective state questionnaire appears to have been sensitive to the environment and produced plausible affective trajectories, there is the possibility that the items are understood by learners in ways that are different than intended. A second possibility is that there were other relevant affective states, such as interest, frustration or boredom, that were not captured by the questionnaire. While it was considered necessary to limit the questionnaire in order to minimise the interruption of practice, a complementary qualitative method, such as interviews, stimulated recall or focus groups, employed after practice sessions could be used to address these issues.

A fourth and final limitation we will mention is concerned with the operationalisation of CF in the GREET platform, which required learners in the CF condition to move to the next question after making a correct attempt and did not provide the opportunity to retry a question. This design removed one possible source of variation that might have been linked to situated affect. Some learners may have re-attempted questions they had already answered correctly to test hypotheses about Dutch language word order or to practice (i.e. automatise) speaking utterances with word orders that were challenging. If learners had been given the choice to re-attempt an answer or not, this would have provided us with the opportunity to observe whether or not learners chose to interact more with the language they produce, which could in turn be used to investigate the potential impact of CF on learner engagement. Adjusting the behaviour of CF in the spoken practice component is trivial, making investigating the relationship between CF and engagement a very feasible possibility.

6.7 Future directions

Quantitative approaches to describing learning processes provide a detailed description of learner and system behaviour over time and allow us to describe precisely the relationship between behaviour and affective variables. However, during the course of completing the work in this thesis, we have found that identifying important patterns of behaviour can be challenging.

We can think of at least two interesting possibilities for addressing this challenge. A first possibility would be to attempt to identify important affective variables or interaction patterns with a purely data-driven approach. This approach would likely require large amounts of participant data, which could be collected through longitudinal studies. Now that we have a functioning platform providing self-study grammar practice with automatic CF, it would be particularly interesting to work together with educators to extend the materials and integrate the system with long running Dutch courses, as Heift

(2016) has done with L2 learners of German. Large, cumulative collections of system-learner interactions lend themselves well to data mining techniques and constitute a rich resource for different types of investigations in the future. While this approach is not feasible within the scope of one doctoral dissertation, we consider it a promising future direction.

A second interesting possibility would be to add a significant in-depth qualitative component. In the CALL literature, we have seen that qualitative methods can provide insights into the relationships between CALL environments and learners' affective responses to those environments (Colpaert, 2010), document the affective impact of social interactions that take place during learning (J. C. Chen & Brown, 2011), and reveal that features assumed to be motivating by researchers are not always the ones appreciated by learners (Gleason & Suvorov, 2012).

Qualitative techniques could be used to identify features of the practice environment with an important role, which could then be cross-checked with quantitative analysis. It would be interesting to know which features of the environment contribute to its positive effect. It seems possible that the videos contributed, and given the ideal L2 self results above, it is plausible that the word order production exercises was appreciated. Therefore, in future research we plan to complement the mainly quantitative methods developed in this thesis by incorporating qualitative methods for studying learners' affective experiences and ideal L2 selves in relation to spoken CALL practice.

A second, timely future direction that seems highly relevant would be to investigate the impact of CF further. Future work could approach this in two ways. One approach could be to conduct a longitudinal study to rule out the possibility of novelty effects and allow more time to study how the affective trajectories of the NOCF and CF groups develop. A second possibility is related to our suggestion above that learners possibly did not expect automatic CF from the computer environment and that therefore the lack of CF was not a factor in their affective response to the system. This could be investigated with a modified experimental design that 1) elicited learners' expectations and beliefs about the capabilities and limitations of computer technologies for language learning before practice and 2) allowed learners to practice with both NOCF and CF conditions (using two groups with different initial practice conditions to control for order). Such a design might draw learners' attention to the presence and absence of CF, stimulate them to make the comparison between the two and possibly elicit different affective responses.

Turning towards more distant goals, we see two interesting possibilities related to integrating macro- and micro-perspectives on affective dimensions of language learning taking place in CALL environments.

A first idea would be to use the platform we have developed to contribute research on the relationship between CALL L2 learning processes and learners' L2 self-guides. The research so far into the relationship between CALL practice and ideal L2 selves appears to be very limited. To our knowledge, only Gleason and Suvorov (2012) have explored this topic, finding that online

spoken L2 practice had minimal impact on learner's ideal and ought-to L2 selves. One possible explanation for the apparent lack of effects is that ideal and ought-to L2 selves are generally considered to be relatively stable affective characteristics (Dörnyei, 2009a) and it may be difficult for CALL practice to have a strong impact on the self-guides without a specially built motivational intervention component. On the other hand, it appears that more situated affective constructs (e.g. interest, enjoyment, anxiety) can be more sensitive to the learning situation (Chapter 3, Dela Rosa & Eskenazi, 2011). Therefore, we suggest another approach to investigating the relationship between CALL practice and ideal L2 selves. We propose treating possible selves as an independent variable that has the potential to interact with the language learning practice context. In this view, affective experiences and practice behaviour during L2 practice may be influenced by the interaction of learner's L2 self-guides and characteristics of the practice environment.

The platform we have developed has the necessary components to investigate this possibility. For example, we could investigate the apparent link between the positive experience that learners reported in Chapters 3 and 4 and their aspirations to speak an L2 from Chapter 5. One clear way to proceed would be to conduct a follow-up experiment in which learners complete word order practice in different modalities. As we have already developed a version of the word order exercise for the written modality (Drozdova, Cucchiari, & Strik, 2013), the barrier to performing this experiment would be very low. An advantage of conducting this research with our platform would be the additional detailed behaviour logs that would be produced, as traditional large-scale motivational studies often employ survey measures of intended effort (e.g. Papi & Teimouri, 2014) which may differ from the effort learners actually exert.

Our final suggestion for future work is one that is more exploratory and is concerned with investigating how CALL systems could use profiles of the type described in Chapter 5 as a resource in advanced tutorial systems. To date, research has demonstrated that CALL systems that track learner performance and model the learner's linguistic knowledge to adapt important aspects of practice like feedback are feasible (Heift & Schulze, 2007), and in some cases have already been successfully employed in educational settings for over a decade (Heift, 2016). CALL systems that model and adapt to learners' affective characteristics are less frequent, but they do exist (e.g. Heilman et al., 2010). Researchers in related domains have also pursued this topic (Walkington & Bernacki, 2014). A question in this area we would like to pursue is whether or not information about a learner's ideal L2 self could be employed as a resource by technology to provide learners with personally relevant L2 speaking practice.

One approach could be to employ the profiles in technology-based agents acting on the learner's behalf that identify L2 content that is related to learners' ideal and ought-to L2 selves. For example, a system module could aim to provide learners with spoken practice that features situations, vocabulary or topics that are related to hopes the person has for the future. This problem has some parallels with recommender systems, which "collect information on

the preferences of its users for a set of items (e.g., movies, songs, books, ... and e-learning material)” to provide “users with predictions and recommendations of items” (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013, p. 1). However, there are additional complications because a module would also have to identify resources based on motivational relevance and linguistic difficulty. This seems to be a direction for CALL research that has the potential to contribute technology for making personally relevant language practice more available. Reviews of research in recommender systems (e.g. Drachsler et al., 2007; Nunes, 2008; Park, Kim, Choi, & Kim, 2012), technology-assisted creating of language learning materials (e.g. C. Chen, Liou, & Chang, 2006; Heilman, Zhao, Pino, & Eskenazi, 2008; Perez-Beltrachini, Gardent, & Kruszewski, 2012) and technology for personalising learning (e.g. Walkington & Bernacki, 2014) would help to judge the feasibility of this idea.

6.8 Final words

This thesis combined advances in speech technology with developments in L2 learning and motivation theory to develop a comprehensive L2 learning platform, for the purpose of investigating affective dimensions in spoken L2 grammar practice. The knowledge and the multiple resources produced in this research show that computer-based language learning environments are valuable research tools with strong potential to contribute to our understanding of affective factors in L2 learning, as well as pedagogical design.

References

- Aerts, A., Colpaert, J., & Oberhofer, M. (Eds.). (2010, August). *Motivation and beyond, Proceedings CALL 2010*. Antwerp, University of Antwerp. Retrieved from http://uahost.uantwerpen.be/linguapolis/scuati/proceedings_CALL%202010.pdf
- Aghaei Pour, P., Hussain, M. S., AlZoubi, O., D'Mello, S., & Calvo, R. A. (2010, June). The Impact of System Feedback on Learners' Affective and Physiological States. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems: 10th International Conference, ITS 2010 Proceedings* (p. 264-273). Pittsburgh, PA, USA: Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-642-13388-6_31 doi: 10.1007/978-3-642-13388-6_31
- Agudo, J. d. D. M. (2013). An investigation into how EFL learners emotionally respond to teachers' oral corrective feedback. *Colombian Applied Linguistics Journal*, 15(2), 265-278.
- Ainley, M., & Hidi, S. (2014). Interest and Enjoyment. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (p. 205-227). Routledge.
- Ainley, M., Hillman, K., & Hidi, S. (2002). Gender and interest processes in response to literary texts: situational and individual interest. *Learning and Instruction*, 12(4), 411-428.
- Akhtar-Danesh, N., Baumann, A., & Cordingley, L. (2008). Q-Methodology in Nursing Research: A Promising Method for the Study of Subjectivity. *Western Journal of Nursing Research*, 30(6), 759-773. doi: 10.1177/0193945907312979
- Alm, A. (2010). The Social Side of L2 Motivation. In *2010 CALL Conference - Motivation and beyond* (p. 10-13). Retrieved from http://uahost.uantwerpen.be/linguapolis/scuati/proceedings_CALL%202010.pdf
- Al-Shehri, A. S. (2009). Motivation and Vision: The Relation Between the Ideal L2 Self, Imagination and Visual Style. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (p. 164-171). Bristol:

- Multilingual Matters.
- Arnold, J., & Brown, D. (1999). A map of the terrain. In J. Arnold (Ed.), *Affect in Language Learning* (p. 1-24). Cambridge: Cambridge University Press.
- Au, S. Y. (1988). A critical appraisal of Gardner's social-psychological theory of second-language (L2) learning. *Language Learning*, 38(1), 75-99. doi: 10.1111/j.1467-1770.1988.tb00402.x
- Bandura, A. (1978). The Self System in Reciprocal Determinism. *American Psychologist*, 33(4), 344-358. doi: 10.1037/0003-066X.33.4.344
- Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs.
- Bax, S. (2003). CALL—past, present and future. *System*, 31(1), 13-28. doi: [http://dx.doi.org/10.1016/S0346-251X\(02\)00071-4](http://dx.doi.org/10.1016/S0346-251X(02)00071-4)
- Beck, J. (2005). Engagement tracing: using response times to model student disengagement. In *Proceedings of the 2005 conference on Artificial Intelligence in Education* (p. 88-95).
- Blashfield, R. K., & Aldenderfer, M. S. (1988). The methods and problems of cluster analysis. In *Handbook of multivariate experimental psychology* (p. 447-473). Springer.
- Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6), 4-16.
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109-132. doi: <http://dx.doi.org/10.1016/j.knosys.2013.03.012>
- Bodnar, S., Cucchiari, C., Strik, H., & van Hout, R. (2016). Evaluating the motivational impact of CALL systems: current practices and future directions. *Computer Assisted Language Learning*, 29(1), 186-212. doi: 10.1080/09588221.2014.927365
- Bodnar, S., Penning de Vries, B., Cucchiari, C., Strik, H., & van Hout, R. (2013, August). Learners' situated motivation in oral grammar practice with an ASR-enabled CALL system. In *Proceedings of SLaTE 2013* (p. 66-71).
- Boekaerts, M. (1988). Motivated learning: bias in appraisals. *International Journal of Educational Research*, 12(3), 267-280.
- Bohnacker, U. (2006). When Swedes begin to learn German. *Second Language Research*, 22(4), 443-486.
- Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005-2014: Understanding a publication surge and a changing landscape. *System*, 55, 145-157. doi: <http://dx.doi.org/10.1016/j.system.2015.10.006>
- Borgen, F. H., & Barnett, D. C. (1971). Applying Cluster Analysis in Counseling Psychology Research. *Journal of Counseling Psychology*, 34(4), 456-468.
- Bowden, J. A., Starrs, C. D., & Quinn, T. J. (1989). Modern Language Teaching in Australian Universities. *Higher Education Research & Development*,

- 8(2), 129-146. doi: 10.1080/0729436890080202
- Brown, D. N. (2010). Potential Pedagogical Benefits of eTandem in Language Learning: Motivation, Self-esteem and Language Knowledge. In *2010 CALL Conference - Motivation and beyond* (p. 48-50). Retrieved from http://uahost.uantwerpen.be/linguapolis/scuati/proceedings_CALL%202010.pdf
- Brown, H. (1981). Affective factors in second language learning. In J. Alatis, H. Altman, & P. Alatis (Eds.), *The Second Language Classroom: Directions for the eighties* (p. 111-129). New York, NY: Oxford University Press.
- Brown, H. D. (2000). *Principles of Language Learning and Teaching*. New York: Longman.
- Brown, S. R. (1980). *Political subjectivity: applications of Q methodology in political science*. New Haven: Yale University Press.
- Brown, S. R. (1993). A primer on Q methodology. *Operant subjectivity*, 16(3/4), 91-138.
- Bush, M. D. (2008). Computer-Assisted Language Learning: From Vision to Reality? *CALICO Journal*, 25(3), 443-470.
- Busse, V., & Walter, C. (2013). Foreign Language Learning Motivation in Higher Education: A Longitudinal Study of Motivational Changes and Their Causes. *The Modern Language Journal*, 97(2), 435-456.
- Bygate, M. (1987). *Speaking* (C. N. Candlin & H. G. Widdowson, Eds.). Oxford University Press.
- CardSorting.net. (2016, May). Retrieved from <http://www.cardsorting.net>
- Carless, D. R. (2003). Factors in the implementation of task-based teaching in primary schools. *System*, 31(4), 485-500. doi: 10.1016/j.system.2003.03.002
- Caron, A., & Debeurme, G. (2010). Distance Learning Courses and Motivation. In *2010 CALL Conference - Motivation and beyond*. doi: http://uahost.uantwerpen.be/linguapolis/scuati/proceedings_CALL%202010.pdf
- Carver, C. S. (2003). Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition and Emotion*, 17(2), 241-261.
- Carver, C. S. (2006). Approach, Avoidance, and the Self-Regulation of Affect and Action. *Motivation and Emotion*, 30(2), 105-110. doi: 10.1007/s11031-006-9044-7
- Cathcart, R. L., & Olsen, J. W. B. (1976). Teachers' and students' preferences for correction of classroom conversation errors. In J. F. Fanselow & R. H. Crymes (Eds.), *On TESOL '76* (p. 41-45). TESOL.
- Chambers, A., & Bax, S. (2006). Making CALL work: Towards normalisation. *System*, 34(4), 465-479. doi: <http://dx.doi.org/10.1016/j.system.2006.08.001>
- Chan, J. C. Y., & Lam, S.-f. (2010). Effects of different evaluative feedback on students' self-efficacy in learning. *Instructional Science*, 38(1), 37-58. doi: 10.1007/s11251-008-9077-2

- Chang, L. (2007). The effects of using CALL on advanced Chinese foreign language learners. *CALICO Journal*, 24(2), 331-353.
- Chang, M. (2005). Applying Self-Regulated Learning Strategies in a Web-Based Instruction - An Investigation of Motivation Perception. *Computer Assisted Language Learning*, 18(3), 217-230. doi: 10.1080/09588220500178939
- Chang, M.-M. (2010). Effects of Self-Monitoring on Web-Based Language Learner's Performance and Motivation. *CALICO Journal*, 27(2), 298-310.
- Chapelle, C. A. (2001). *Computer Applications in Second Language Acquisition*. Cambridge University Press: Cambridge, UK.
- Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning*, 27(1), 29-46. doi: 10.1111/j.1467-1770.1977.tb00290.x
- Chen, C., Liou, H., & Chang, J. S. (2006). FAST: An Automatic Generation System for Grammar Tests. In *Proceedings of the COLING/ACL on Interactive presentation sessions* (p. 1-4).
- Chen, J. C., & Brown, K. L. (2011). The effects of authentic audience on English as a second language (ESL) writers: a task-based, computer-mediated approach. *Computer Assisted Language Learning*, 25(5), 1-20. doi: 10.1080/09588221.2011.606224
- Chenoweth, N. A., Day, R. R., Chun, A. E., & Luppescu, S. (1983). Attitudes and Preferences of ESL Students to Error Correction. *Studies in Second Language Acquisition*, 6(1), 79-87.
- Chiu, T.-L., Liou, H.-C., & Yeh, Y. (2007). A Study of web-based oral activities enhanced by Automatic Speech Recognition for EFL college learning. *Computer Assisted Language Learning*, 20(3), 209-233. doi: 10.1080/09588220701489374
- Choi, S., & Johnson, W. (2005). Assessing Dynamic Aspects of Learner Motivation in Simulation/Gaming Based Foreign Language Learning Environment. In *AIED 2005 - Workshop on motivation and affect in educational software*.
- Clément, R. (1986). Second Language Proficiency and Acculturation: An Investigation of the Effects of Language Status and Individual Characteristics. *Journal of Language and Social Psychology*, 5(4), 271-290. doi: 10.1177/0261927X8600500403
- Clifford, R., & Granoien, N. (2008). Applications of Technology to Language Acquisition Processes. In V. M. Holland & F. P. Fisher (Eds.), *The Path of Speech Technologies in Computer Assisted Language Learning: From Research Toward Practice* (p. 25-43). Routledge : New York.
- Colpaert, J. (2010). Elicitation of language learners' personal goals as design concepts. *Innovation in Language Learning and Teaching*, 4(3), 259-274.
- Coniam, D. (1998). The Use of Speech Recognition Software as an English Language Oral Assessment Instrument: An Exploratory Study. *CALICO Journal*, 15(4), 7-23.

- Coniam, D. (1999). Voice Recognition Software Accuracy with Second Language Speakers of English. *System*, 27(1), 49-64.
- Cook, V. (1988). Designing a BASIC Parser for CALL. *CALICO Journal*, 6(1), 50-67.
- Cornillie, F., Clarebout, G., & Desmet, P. (2012). Between learning and playing? Exploring learners perceptions of corrective feedback in an immersive game for English pragmatics. *ReCALL*, 24(3), 257-278. doi: 10.1017/S0958344012000146
- Cornillie, F., & Desmet, P. (2013). Seeking out fun failure: how positive failure feedback could enhance the instructional effectiveness of CALL mini-game. In *Global perspectives on Computer-Assisted Language Learning. Proceedings of WorldCALL 2013* (p. 64-68).
- Cornillie, F., Vandewaetere, M., & Jordaens, K. (2010). Wedding Games with Instructional Design: an Impossible Match? In *2010 CALL Conference - Motivation and beyond*. Retrieved from <http://uahost.uantwerpen.be/linguapolis/scuati/proceedings.CALL%202010.pdf>
- Crookes, G., & Arakaki, L. (1999). Teaching idea sources and work conditions in an ESL program. *Tesol Journal*, 8(1), 15-19.
- Crookes, G., & Schmidt, R. W. (1991). Motivation: Reopening the Research Agenda. *Language Learning*, 41(4), 469-512. doi: 10.1191/1362168806lr206pr
- Csikszentmihalyi, M. (1975). A Theoretical Model for Enjoyment. In *Beyond boredom and anxiety* (p. 35-54). San Francisco: San Francisco: Jossey-Bass.
- Csizér, K., & Dörnyei, Z. (2005a). The internal structure of language learning motivation and its relationship with language choice and learning effort. *The Modern Language Journal*, 89(1), 19-36. doi: 10.1111/j.0026-7902.2005.00263.x
- Csizér, K., & Dörnyei, Z. (2005b). Language Learners Motivational Profiles and Their Motivated Learning Behavior. *Language Learning*, 55(4), 613-659. doi: 10.1111/j.0023-8333.2005.00319.x
- Csizér, K., & Kormos, J. (2009). Learning experiences, selves and motivated learning behaviour : a comparative analysis of structural models for Hungarian secondary and university learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (p. 98-117). Multilingual Matters.
- Csizér, K., & Lukács, G. (2010). The comparative analysis of motivation, attitudes and selves: The case of English and German in Hungary. *System*, 38(1), 1-13. doi: <http://dx.doi.org/10.1016/j.system.2009.12.001>
- Cucchiari, C., Doremalen, J., & Strik, H. (2012). Practice and feedback in L2 speaking: an evaluation of the DISCO CALL system. In *INTERSPEECH-2012* (p. 779-782).
- Cucchiari, C., Neri, A., & Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, 51(10), 853-863. doi: <http://dx.doi.org/10.1016/j.specom.2009>

.03.003

- Dasse-Askildson, V. (2008). How Learners Affective Variables Impact Their Perception Of Recasts In The Acquisition Of Grammatical Gender In L2 French. *Arizona Working Papers in SLA & Teaching*, 15, 1-35.
- de Bot, K. (1996). The Psycholinguistics of the Output Hypothesis. *Language Learning*, 46(3), 529-555. doi: 10.1111/j.1467-1770.1996.tb01246.x
- de Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7-21. doi: 10.1017/S1366728906002732
- de Vicente, A., & Pain, H. (1999). Motivation self-report in ITS. In *In Proceedings of the 9th World Conference on Artificial Intelligence in Education* (p. 648-650).
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer.
- DeKeyser, R. (2007). Introduction: Situating the concept of practice. In R. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (p. 1-19). Cambridge, USA: Cambridge University Press.
- del Soldato, T., & du Boulay, B. (1995). Implementation of motivational tactics in tutoring systems. *International Journal of Artificial Intelligence in Education*, 6(4), 337-378.
- Dela Rosa, K., & Eskenazi, M. (2011). Self-assessment of Motivation: Explicit and Implicit Indicators in L2 Vocabulary Learning. In B. S. K. J. Biswas G. & A. Mitrovic (Eds.), *Artificial Intelligence in Education* (p. 296-303). Springer Berlin Heidelberg.
- Dembovskaia, S. B. (2010). Motivational and Strategic Pre-tasks in a Foreign Language Class. In *2010 CALL Conference - Motivation and beyond*. Retrieved from http://uahost.uantwerpen.be/linguapolis/scuati/proceedings_CALL%202010.pdf
- Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does Popular Speech Recognition Software Work with ESL Speech? *TESOL Quarterly*, 34(3), 592-603.
- Develotte, C., Guichona, N., & Vincenta, C. (2010). The use of the webcam for teaching a foreign language in a desktop videoconferencing environment, *ReCALL*. *ReCALL*, 22(3), 293-312. doi: doi:10.1017/S0958344010000170
- de Vicente, A., & Pain, H. (2002). Informing the Detection of the Students' Motivational State: An Empirical Study. In *Proceedings of the 6th International Conference on Intelligent Tutoring Systems* (p. 933-943). London, UK, UK: Springer-Verlag.
- Dewaele, J.-M. (2009). Perception, Attitude and Motivation. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *Language Teaching and Learning* (p. 163-192). London, UK: Continuum.
- Dewaele, J.-M. (2010). *Emotions in Multiple Languages* (First Edition ed.). Palgrave Macmillan.

- Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, 4(2), 237-274.
- Dörnyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British Journal of Educational Psychology*, 70(4), 519-538.
- Dörnyei, Z. (2002). The motivational basis of language learning tasks. In P. Robinson (Ed.), *Individual differences and instructed language learning* (p. 137-158). Amsterdam: John Benjamins.
- Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2005). Motivation and 'Self-Motivation'. In *The Psychology Of The Language Learner: Individual Differences In Second Language Acquisition* (p. 93-108). Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2009a). The L2 Motivational Self System. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (p. 9-42). Multilingual Matters.
- Dörnyei, Z. (2009b). *The psychology of second language acquisition*. Oxford University Press.
- Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, Language Attitudes And Globalisation: A Hungarian Perspective*. Multilingual Matters.
- Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics*, 4, 43-69.
- Dörnyei, Z., & Ushioda, E. (Eds.). (2009). *Motivation, Language Identity and the L2 Self*. Bristol, England: Multilingual Matters.
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation (2nd ed.)* (C. N. Candlin & D. R. Hall, Eds.). Harlow: Longman.
- Drachler, H., Hummel, H., Van den Berg, B., Eshuis, J., Berlanga, A., Nadolski, R., ... Koper, R. (2007). Recommendation strategies for e-learning: preliminary effects of a personal recommender system for lifelong learners.. Retrieved 2016-01-21, from <http://lnx-hr1-075v.web.pwo.ou.nl/handle/1820/1010>
- Drozdova, P., Cucchiari, C., & Strik, H. (2013, August). L2 syntax acquisition: the effect of oral and written computer assisted practice. In *INTERSPEECH 2013, 14th annual conference of the international speech communication association* (p. 982-986). Lyon, France.
- Dunkel, C., & Kerpelman, J. (Eds.). (2006). *Possible selves : theory, research and applications*. Nova Science Publishers.
- Dunnette, M. D., Campbell, J. P., & Hakel, M. D. (1967). Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Organizational Behavior and Human Performance*, 2(2), 143-174. doi: [http://dx.doi.org/10.1016/0030-5073\(67\)90027-X](http://dx.doi.org/10.1016/0030-5073(67)90027-X)
- Dziopa, F., & Ahern, K. (2011). A Systematic Literature Review of the Applications of Q-Technique and Its Methodology. *Methodology*, 7(2), 39-55. doi: 10.1027/1614-2241/a000021

- Egbert, J. (2003). A Study of Flow Theory in the Foreign Language Classroom. *The Modern Language Journal*, 87(4), 499-518.
- Ehsani, F., & Knodt, E. (1998). Speech technology in Computer-Aided Language Learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology*, 2(1), 45-60.
- Ellis, R. (1985). *Understanding Second Language Acquisition*. Oxford Oxfordshire: Oxford University Press.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford University Press.
- Ellis, R. (2009). Corrective Feedback and Teacher Development. *L2 Journal*, 1(1), 3-18.
- Ellis, R. (2010). A Framework for Investigating Oral and Written Corrective Feedback. *Studies in Second Language Acquisition*, 32(2), 335-349.
- Erlam, R., & Loewen, S. (2010). Implicit and Explicit Recasts in L2 Oral French Interaction. *Canadian Modern Language Review*, 66(6), 877-905. doi: <http://dx.doi.org/10.3138/cmlr.66.6.877>
- Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication*, 51(10), 832-844. doi: <http://dx.doi.org/10.1016/j.specom.2009.04.005>
- Felix, U. (2008). The unreasonable effectiveness of CALL: What have we learned in two decades of research? *ReCALL*, 20(2), 141-161. doi: 10.1017/S0958344008000323
- Ferreira da Silva, P., García, M. R., Edwards, P., & Domínguez-Noriega, S. (2010). Lets Talk Together: Developing Communicative Competence through Avatar Interaction in Virtual Worlds. In *2010 CALL Conference - Motivation and beyond*. Retrieved from http://uahost.uantwerpen.be/linguapolis/scuati/proceedings_CALL%202010.pdf
- Fidaoui, D., Bahousa, R., & Bacha, N. N. (2010). CALL in Lebanese elementary ESL writing classrooms. *Computer Assisted Language Learning*, 23(2), 151-168. doi: 10.1080/09588221003666248
- Fischer, R. (2007). How do we Know what Students are Actually Doing? Monitoring Students' Behavior in CALL. *Computer Assisted Language Learning*, 20(5), 409-442. doi: 10.1080/09588220701746013
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), 59-109. doi: 10.3102/00346543074001059
- Fredrickson, B. L. (2001). The Role of Positive Emotions in Positive Psychology: The Broaden-and-Build Theory of Positive Emotions. *The American Psychologist*, 56(3), 218-226.
- Gamper, J., & Knapp, J. (2002). A review of intelligent CALL systems. *Computer Assisted Language Learning*, 15(4), 329-342.
- Gardner, R. C. (1985a). *The Attitude/Motivation Test Battery: Technical Report* (Tech. Rep.). University of Western Ontario.
- Gardner, R. C. (1985b). *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London, UK: Edward Arnold.

- Gardner, R. C., Day, B., & MacIntyre, P. (1992). Integrative Motivation, Induced Anxiety, and Language Learning in a Controlled Environment. *Studies in Second Language Acquisition*, 14(2), 197-214.
- Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second language. *Canadian Journal of Psychology*, 13(4), 266-272.
- Gardner, R. C., & MacIntyre, P. D. (1993). A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching*, 26(1), 1. doi: 10.1017/S0261444800000045
- Garrett, P., & Young, R. F. (2009). Theorizing Affect in Foreign Language Learning: An Analysis of One Learner's Responses to a Communicative Portuguese Course. *The Modern Language Journal*, 93(2), 209-226. doi: 10.1111/j.1540-4781.2009.00857.x
- Gillet, N., Vallerand, R. J., Lafrenière, M.-A. K., & Bureau, J. S. (2012). The mediating role of positive and negative affect in the situational motivation-performance relationship. *Motivation and Emotion*, 37(3), 465-479. doi: 10.1007/s11031-012-9314-5
- Gleason, J., & Suvorov, R. (2012). Learner Perceptions of Asynchronous Oral Computer-mediated Communication: Proficiency and Second Language Selves. *The Canadian Journal of Applied Linguistics*, 15(1), 100-121.
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richards, D. L., & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70-105. doi: 10.1080/09588221.2012.700315
- Hartigan, J. A., & Hartigan, P. M. (1985). The Dip Test of Unimodality. *The Annals of Statistics*, 13(1), 70-84.
- Heift, T. (2003). Drag or Type, But Dont Click: A Study on the Effectiveness of Different CALL Exercise Types. *Canadian Journal of Applied Linguistics*, 6(1), 69-85.
- Heift, T. (2008). Modeling learner variability in CALL. *Computer Assisted Language Learning*, 24(4), 305-321. doi: doi:10.1080/09588220802343421
- Heift, T. (2013). Second Language Interaction in Diverse Educational Contexts. In K. McDonough & A. Mackey (Eds.), (p. 189-208). John Benjamins Publishing Company. doi: DOI:10.1075/llt.34.14ch10
- Heift, T. (2016). Web Delivery of Adaptive and Interactive Language Tutoring: Revisited. *International Journal of Artificial Intelligence in Education*, 26(1), 489-503.
- Heift, T., & Chapelle, C. A. (2012). The Routledge Handbook of Second Language Acquisition. In A. Mackey & S. M. Gass (Eds.), *Language learning through technology* (p. 555-569). Routledge : New York.
- Heift, T., & Schulze, M. (2007). *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues*. Routledge.
- Heift, T., & Schulze, M. (2015). Tutorial computer-assisted language learning. *Language Teaching*, 48(4), 471-490. doi: 10.1017/S0261444815000245
- Heilman, M., Collins-Thompson, K., Eskenazi, M., Juffs, A., & Wilson,

- L. (2010). Personalization of Reading Passages Improves Vocabulary Acquisition. *International Journal of Artificial Intelligence in Education*, 20(1), 73-98.
- Heilman, M., Zhao, L., Pino, J., & Eskenazi, M. (2008). Retrieval of Reading Materials for Vocabulary and Reading Practice. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications* (p. 80-88). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Herron, D., & Menzel, W. (2000, March). Error diagnosis for spoken language ISLE Deliverable Deliverable 4.5 [Computer software manual]. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.161.8530>
- Hessel, G. (2015). From vision to action: Inquiring into the conditions for the motivational capacity of ideal second language selves. *System*, 52, 103-114. doi: 10.1016/j.system.2015.05.008
- Higgins, E. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94, 319-340.
- Higgins, E., Klein, R., & Strauman, T. (1985). Self-concept discrepancy theory: A psychological model for distinguishing among different aspects of depression and anxiety. *Social Cognition*, 3(1), 51-76.
- Holland, V. M., & Fisher, F. P. (Eds.). (2007). *The Path of Speech Technologies in Computer Assisted Language Learning: From Research Toward Practice*. Routledge Studies in Computer Assisted Language Learning. Oxford: Routledge.
- Horwitz, E. K. (2010). Foreign and second language anxiety. *Language Teaching*, 43(2), 154-167.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132.
- Hubbard, P. (2005). A review of subject characteristics in CALL research. *Computer Assisted Language Learning*, 18(5), 351-368.
- Irie, K. (2014). Q methodology for post-social-turn research in SLA. *Studies in Second Language Learning and Teaching*, 4(1), 13-32. doi: 10.14746/ssllt.2014.4.1.2
- Irie, K., & Ryan, S. (2015). Study abroad and the dynamics of change in learner L2 self-concept. In *Motivational dynamics in language learning* (p. 343-366). Bristol, UK: Multilingual Matters.
- Islam, M., Lamb, M., & Chambers, G. (2013). The L2 Motivational Self System and National Interest: A Pakistani perspective. *System*, 41(2), 231-244. doi: <http://dx.doi.org/10.1016/j.system.2013.01.025>
- Izquierdo, J., Simard, D., & Pulido, M. G. G. (2015). Multimedia Instruction & Language Learning Attitudes: A Study with University Students. *Revista Electrónica de Investigación Educativa*, 17(2), 101-115.
- Jamieson, J., & Chapelle, C. (1988). Using CALL Effectively: What Do We Need to Know about Students? *System*, 16(2), 151-62.
- Jauregi, K., de Graaff, R., van den Bergh, H., & Kriz, M. (2012). Native/non-

- native speaker interactions through video-web communication: a clue for enhancing motivation? *Computer Assisted Language Learning*, 25(1), 1-19. doi: 10.1080/09588221.2011.582587
- Jean, G., & Simard, D. (2011). Grammar Teaching and Learning in L2: Necessary, but Boring? *Foreign Language Annals*, 44(3), 467-494.
- Johnson, W. L. (2010). Serious Use of a Serious Game for Language Learning. *International Journal of Artificial Intelligence in Education*, 20, 175-195. doi: doi:10.3233/JAI-2010-0006
- Johnson, W. L., & Valente, A. (2009). Tactical Language and Culture Training Systems: Using AI to Teach Foreign Languages and Cultures. *AI Magazine*, 30(2), 72-83.
- Jordens, P. (1988). The acquisition of word order in Dutch and German as L1 and L2. *Second Language Research*, 4(1), 41-65.
- Julkunen, K. (2001). Situation- and task-specific motivation in foreign language learning. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (p. 29 - 41). University of Hawaii Press.
- Kartchava, E. (2012). *Noticeability of corrective feedback, L2 development and learner beliefs* (Unpublished doctoral dissertation). Université de Montréal.
- Kartchava, E., & Ammar, A. (2013). Learners Beliefs as Mediators of What Is Noticed and Learned in the Language Classroom. *TESOL Quarterly*, 48(1), 86109.
- Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, 10(3), 2-9.
- Kerpelman, J. L. (2006). Using Q Methodology to study possible selves. In C. Dunkel & J. Kerpelman (Eds.), *Possible Selves: Theory, Research and Application* (p. 164-186). Nova.
- Kim, T.-Y. (2012). The L2 Motivational Self System of Korean EFL Students: Cross-grade Survey Analysis. *English Teaching*, 67(1), 29-51.
- King, E. (2011). Possibility Spaces: Using The Sims 2 as a Sandbox to Explore Possible Selves with At-Risk Teenage Males. *International Journal of Game-Based Learning*, 1(2), 34-51.
- Kissau, S. (2006). Gender differences in second language motivation: An investigation of micro- and macro-level influences. *Canadian Journal of Applied Linguistics*, 9(1), 73-96.
- Kissau, S. (2012). Perceptions of self-efficacy for two types of second language methods instruction. *Computer Assisted Language Learning*, 20(4), 295-317. doi: 10.1080/09588221.2011.587436
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press Inc.
- Lamb, M. (2013). 'Your mum and dad can't teach you!': constraints on agency among rural learners of English in the developing world. *Journal of Multilingual and Multicultural Development*, 34(1), 14-29. doi: 10.1080/01434632.2012.697467
- Lasagabaster, D., & Sierra, J. M. (2005). Error Correction: Students' versus

- Teachers' Perceptions. Language Awareness. *Language Awareness*, 14(2-3), 112-127.
- Lehtonen, T., & Tuomainen, S. (2003). CSCL - a tool to motivate foreign language learners: The Finnish application. *ReCALL*, 15(1), 51-67. doi: 10.1017/S095834400300051X
- Levy, M., Hubbard, P., Stockwell, G., & Colpaert, J. (2015). Research challenges in CALL. *Computer Assisted Language Learning*, 28(1), 1-6. Retrieved from <http://dx.doi.org/10.1080/09588221.2014.987035> doi: 10.1080/09588221.2014.987035
- Loewen, S., Li, S., Fei, F., Thompson, A., Nakatsukawas, K., Ahn, S., & Chen, X. (2009). Second Language Learners Beliefs About Grammar Instruction and Error Correction. *The Modern Language Journal*, 93(1), 91-104.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19(1), 37-66.
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46(1), 1-40.
- MacIntyre, P., & Serroul, A. (2015). Motivation on a Per-Second Timescale: Examining Approach-Avoidance Motivation During L2 Task Performance. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational Dynamics in Language Learning* (p. 109-138). Multilingual Matters.
- MacIntyre, P. D. (2002). Motivation, anxiety and emotion in second language acquisition. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (p. 45-69). Amsterdam / Philadelphia: John Benjamins.
- Macintyre, P. D. (2007). Willingness to Communicate in the Second Language: Understanding the Decision to Speak as a Volitional Process. *The Modern Language Journal*, 91(4), 564-576.
- MacIntyre, P. D., Burns, C., & Jessome, A. (2011). Ambivalence About Communicating in a Second Language: A Qualitative Study of French Immersion Students Willingness to Communicate. *The Modern Language Journal*, 95(1), 81-96.
- MacIntyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. (1998). Conceptualizing Willingness to Communicate in a L2: A Situational Model of L2 Confidence and Affiliation. *The Modern Language Journal*, 82(4), 545-562.
- MacIntyre, P. D., & Gregersen, T. (2012). Affect: The Role of Language Anxiety and Other Emotions in Language Learning. In S. R. Sarah Mercer & M. Williams (Eds.), *Psychology for Language Learning - Insights from Research, Theory and Practice* (p. 103-118). Palgrave Macmillan.
- MacIntyre, P. D., & Legatto, J. J. (2011). A Dynamic System Approach to Willingness to Communicate: Developing an Idiodynamic Method to Capture Rapidly Changing Affect. *Applied Linguistics*, 32(2), 149-171. doi: 10.1093/applin/amq037
- MacIntyre, P. D., Mackinnon, S. P., & Clement, R. (2009). Possible Selves as a source of language learning motivation. *Motivation, Language Identity*

- and the L2 Self, 193-214.
- MacIntyre, P. D., Mackinnon, S. P., & Clément, R. (2009). Toward the development of a scale to assess possible selves as a source of language learning motivation. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (p. 193-214). Multilingual Matters.
- Markus, H., & Nurius, P. (1986). Possible Selves. *American Psychologist*, 41(9), 954-969.
- Monterrat, B., Desmarais, M., Lavoué, É., & George, S. (2015, June). A Player Model for Adaptive Gamification in Learning Environments. In C. Conati, N. Heffernan, A. Mitrovic, & F. M. Verdejo (Eds.), *Artificial Intelligence in Education: 17th International Conference, AIED 2015* (p. 297-306). Madrid, Spain: Springer International Publishing. doi: 10.1007/978-3-319-19773-9_30
- Moos, D., & Azevedo, R. (2008). Exploring the fluctuation of motivation and use of self-regulatory processes during learning with hypermedia. *Instructional Science*, 36(3), 203-231.
- Morton, H., & Jack, M. (2010). Speech interactive computer-assisted language learning: a cross-cultural evaluation. *Computer Assisted Language Learning*, 23(4), 295-319 (25).
- Motteram, G., & Sharma, P. (2009). Blending Learning in a Web 2.0 World. *International Journal of Emerging Technologies & Society*, 7(2), 83-96.
- Narciss, S. (2004). The Impact of Informative Tutoring Feedback and Self-Efficacy on Motivation and Achievement in Concept Learning. *Experimental Psychology*, 51(3), 214-228. doi: 10.1027/1618-3169.51.3.214
- Neri, A. (2007). *Pedagogical effectiveness of ASR-based reader assisted pronunciation training* (Unpublished doctoral dissertation). University Nijmegen.
- Neri, A., Cucchiaroni, C., & Strik, W. (2003). Automatic Speech Recognition for second language learning: How and why it actually works. In *In Proceeding of International Congresses of Phonetic Sciences* (p. 1157-1160).
- Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21(5), 393-408.
- Noels, K. (2009). The internalization of language learning into the self and social identity. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (p. 295-313). Multilingual Matters.
- Nunes, M. A. S. N. (2008). *Recommender Systems based on Personality Traits* (Unpublished doctoral dissertation). Université Montpellier II.
- Ockwell, D. G. (2008). 'Opening up' Policy to Reflexive Appraisal: A Role for Q Methodology? A Case Study of Fire Management in Cape York, Australia. *Policy Sciences*, 41(4), 263-292.
- OptimalSort. (2016, February). Retrieved from <https://www.optimalworkshop.com/optimalsort>
- Öztürk, G., & Gürbüz, N. (2014). Speaking anxiety among Turkish EFL

- learners: The case at a state university. *Journal of Language and Linguistic Studies*, 10(1), 1-17.
- Packard, B. W.-L., & Conway, P. F. (2006). Methodological Choice and Its Consequences for Possible Selves Research. *Identity*, 6(3), 251-271. doi: 10.1207/s1532706xld0603_3
- Papi, M., & Teimouri, Y. (2014). Language Learner Motivational Types: A Cluster Analysis Study. *Language Learning*, 64(3), 493-525. doi: 10.1111/lang.12065
- Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11), 10059 - 10072. doi: <http://dx.doi.org/10.1016/j.eswa.2012.02.038>
- Payne, J. S., & Whitney, P. J. (2002). Developing L2 Oral Proficiency through Synchronous CMC: Output, Working Memory, and Interlanguage Development. *CALICO Journal*, 20(1), 7-32.
- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wiley (Eds.), *Handbook of research on student engagement* (p. 259-282). New York, NY: Springer.
- Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to Emotions in Education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (p. 1-10). New York, NY: Taylor & Francis.
- Pemberton, & Cook. (2012). Self-directed learning: Concepts, practice, and a novel research methodology. In S. M. Mercer, S. Ryan, & M. Williams (Eds.), *Psychology for language learning: Insights from research, theory, and practice* (p. 305 - 327). Basingstoke: Palgrave Macmillan.
- Penning de Vries, B. (2015). *Computerised Speaking Practice: The Role of Automatic Corrective Feedback in Learning L2 Grammar* (Unpublished doctoral dissertation). Radboud Universiteit Nijmegen.
- Penning de Vries, B., Cucchiari, C., Bodnar, S., Strik, H., & van Hout, R. (2014a). Effect of corrective feedback for learning verb second. *International Review of Applied Linguistics in Language Teaching*.
- Penning de Vries, B., Cucchiari, C., Bodnar, S., Strik, H., & van Hout, R. (2014b). Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning*, 28(6), 550-576.
- Penning de Vries, B., Cucchiari, C., Bodnar, S., Strik, H., & van Hout, R. (in press). Effect of corrective feedback for learning verb second. *International Review of Applied Linguistics in Language Teaching*. *Effect of corrective feedback for learning verb second. International Review of Applied Linguistics in Language Teaching*.
- Perez-Beltrachini, L., Gardent, C., & Kruszewski, G. (2012, June). Generating Grammar Exercises. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP* (p. 147-156). Montréal, Canada: Association for Computational Linguistics. Retrieved from <http://www>

- .aclweb.org/anthology/W12-2017
- Plonsky, L., & Mills, S. (2006). An exploratory study of differing perceptions of error correction between teacher and students: Bridging the gap. *Applied Language Learning*, 16(1), 55-74.
- Presson, N., Davy, C., & MacWhinney, B. (2013). Experimentalized CALL for adult second language learners. In J. W. Schwieter (Ed.), *Innovative Research and Practices in Second Language Acquisition and Bilingualism* (p. 139-164). John Benjamins Publishing Company.
- Qian, K., & McCormick, R. (2014). Building course cohesion: the use of online forums in distance Chinese language learning. *Computer Assisted Language Learning*, 27(1), 44-69. doi: doi:10.1080/09588221.2012.695739
- Raby, F. (2007). A triangular approach to motivation in Computer Assisted Autonomous Language Learning (CAALL). *ReCALL*, 19(2), 21-38. doi: 10.1017/S0958344007000626
- Raby, F. (2010). Computer Assisted Language Learning Motivation (CALLM): a Polytopic, Weighted and Ergonomic Model. In *2010 CALL Conference - Motivation and beyond* (p. 29-31). Retrieved from http://uahost.uantwerpen.be/linguapolis/scuati/proceedings_CALL%202010.pdf
- Raimes, A. (1991). Errors: Windows into the mind. *College ESL*, 1(2), 55-64.
- Sagae, A., Johnson, W. L., & Bodnar, S. (2010). Validation of a Dialog System for Language Learners. In *SIGDIAL 2010* (p. 241-244).
- Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition: A Series of Empirical Studies* (p. 229-248). Oxford: Oxford University Press N.
- Sagarra, N., & Abbuhl, R. (2013). Optimizing the Noticing of Recasts via Computer-Delivered Feedback: Evidence That Oral Input Enhancement and Working Memory Help Second Language Learning. *The Modern Language Journal*, 97(1), 196-216. doi: 10.1111/j.1540-4781.2013.01427.x
- Sauro, S. (2012). L2 performance in text-chat and spoken discourse. *System*, 40(3), 335-348. doi: 10.1016/j.system.2012.08.001.
- Schulz, R. A. (2001). Cultural Differences in Student and Teacher Perceptions concerning the Role of Grammar Instruction and Corrective Feedback: USA: Colombia. *The Modern Language Journal*, 85(2), 244-258.
- Schulze, M. (2003). Grammatical Errors and Feedback Some Theoretical Insights. *CALICO Journal*, 20(3), 437-450. doi: 10.1558/cj.v20i3.
- Sheen, Y. (2008). Recasts, Language Anxiety, Modified Output, and L2 Learning. *Language Learning*, 58(4), 835-874.
- Sheen, Y., & Lyster, R. (2010). The role of oral and written corrective feedback in SLA. *Studies in Second Language Acquisition*, 32(2), 181-201.
- Shepard, B., & Marshall, A. (1999). Possible selves mapping: Life-career

- exploration with young adolescents. *Canadian Journal of Counselling*, 33(1), 37-54.
- Shuman, V., & Scherer, K. R. (2014). Concepts and structures of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (p. 13 - 35). New York, NY: Taylor & Francis.
- Smith, B. (2013). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning & Technology*, 16(3), 53-81.
- Strik, H., Cornillie, F., Colpaert, J., Doremalen, J. V., & Cucchiarini, C. (2009). Developing a CALL System for Practicing Oral Proficiency: How to Design for Speech Technology, Pedagogy and Learners. In *Proceedings of the SLaTE-2009 workshop*. Warwickshire, England.
- Strik, H., Neri, A., & Cucchiarini, C. (2008, February). Speech technology for language tutoring. In *Proceedings of LangTech-2008*. Rome.
- Strik, H., van Doremalen, J., Colpaert, J., & Cucchiarini, C. (2013). Development and Integration of Speech technology into COurseware for language learning: the DISCO project. In P. S. . J. Odijk (Ed.), *Essential Speech and Language Technology for Dutch - Results by the STEVIN-programme* (p. 323-338). Berlin: Springer.
- Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: a comparative study. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (p. 66-97). Bristol: Multilingual Matters.
- Taylor, F. (2013). *Self and identity in adolescent foreign language learning* (D. Singleton, Ed.). Multilingual Matters.
- Tennant, J., & Gardner, R. (2004). The Computerized Mini-AMTB. *CALICO Journal*, 21(2), 245-263.
- Thang, S. M., & Bidmeshki, L. (2010). Investigating the perceptions of UKM undergraduates towards an English for science and technology online course. *Computer Assisted Language Learning*, 23 (1), 1-20.
- Thompson, A. S., & Vásquez, C. (2015). Exploring Motivational Profiles Through Language Learning Narratives. *The Modern Language Journal*, 99(1), 158-174. doi: 10.1111/modl.12187
- Tremblay, P. F., Goldberg, M. P., & Gardner, R. C. (1995). Trait and state motivation and the acquisition of Hebrew vocabulary. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 27(3), 356-370.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327-69.
- Truscott, J. (1999). What's wrong with oral grammar correction. *The Canadian Modern Language Review*, 55(4), 437-456.
- Turner, J. C., & Trucano, M. (2014). Measuring situated emotion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (p. 643-658). New York, NY: Routledge.
- Ushida, E. (2005). The Role of Students' Attitudes and Motivation in Second Language Learning in Online Language Courses. *CALICO Journal*,

- 23(1), 49-78.
- Ushioda, E. (1993). Redefining motivation from the L2 learner's point of view. *Teanga*, 13, 1-12.
- Ushioda, E. (1994). L2 motivation as a qualitative construct. *Teana*, 14 (76-84).
- Ushioda, E. (1996). Developing a dynamic concept of L2 motivation. In T. Hickey & J. Williams (Eds.), *Language, education and society in a changing world* (p. pp. 239-245). Dublin, Ireland: IRAAL/Multilingual Matters.
- Ushioda, E. (2000). Tandem language learning via e-mail: from motivation to autonomy. *ReCALL*, 12(2), 121-128.
- Ushioda, E. (2007). Motivation, autonomy and sociocultural theory. In P. Benson (Ed.), *Learner Autonomy 8: Teacher and Learner Perspectives* (p. 5-24). Authentik.
- Ushioda, E. (2009a). Motivation, Language Identities and the L2 Self: A Theoretical Overview. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (p. 1-8). Multilingual Matters.
- Ushioda, E. (2009b). A Person-in-Context Relational View of Emergent Motivation, Self, and Identity. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (p. 215-228). Multilingual Matters.
- Ushioda, E. (2011). Language learning motivation, self and identity : current theoretical perspectives. *Computer Assisted Language Learning*, 24(3), 199-210.
- Ushioda, E. (2012). Motivation: L2 Learning as a Special Case? In S. R. Sarah Mercer & M. Williams (Eds.), *Psychology for Language Learning - Insights from Research, Theory and Practice* (p. 58-73). Palgrave Macmillan.
- Ushioda, E., & Dörnyei, Z. (2012). Motivation. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (p. 396-409). New York: Routledge.
- Uzum, B. (2011). Influences of Learners' Motivation on the Processing of Corrective Feedback. *Applied Linguistics Forum*, 31, 1-4.
- Van Aacken, S. (1999). What motivates L2 learners in acquisition of Kanji using CALL: A case study. *Computer Assisted Language Learning*, 12(2), 113-136. doi: 10.1076/call.12.2.113.5723
- van Doremalen, J., Strik, H., & Cucchiari, C. (2010, September). Optimizing non-native speech recognition for CALL applications. In *Proceedings of Interspeech 2009* (p. 592-595). Brighton, U.K..
- Vandewaetere, M., & Desmet, P. (2009). Introducing psychometrical validation of questionnaires in CALL research: the case of measuring attitude towards CALL. *Computer Assisted Language Learning*, 22(4), 349-380. doi: 10.1080/09588220903186547
- Vásquez, C., & Harvey, J. (2010). Raising teachers' awareness about corrective feedback through research replication. *Language Teaching Research*, 14(4), 421-443.

- Vinther, J. (2011). Enhancing motivation with cultural narratives in computer-mediated communication. *Computer Assisted Language Learning*, 24(4), 337-352.
- Volle, L. M. (2005). Analyzing Oral Skills in Voice E-Mail and Online Interviews. *Language Learning & Technology*, 9(3), 146-163.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes* (M. Cole, Ed.). Harvard University Press.
- Wachowicz, K. A., & Scott, B. (1999). Software That Listens: It's Not a Question of Whether, It's Question of How. *CALICO Journal*, 16(3), 253-276.
- Walkington, C., & Bernacki, M. L. (2014). Motivating Students by Personalizing Learning around Individual Interests: A Consideration of Theory, Design, and Implementation Issues. In S. A. Karabenick & T. C. Urdan (Eds.), *Motivational Interventions* (Vol. 18, p. 139-176). Emerald Group Publishing Limited.
- Wang, H., Chignell, M., & Ishizuka, M. (2006). Emphatic Tutoring Software Agents Using Real-time Eye Tracking. In *Proceedings of the 2006 symposium on Eye tracking research & applications*. Retrieved from <http://www.miv.t.u-tokyo.ac.jp/papers/wang-EyeTracking-ETRA06.pdf>
- Wang, H., Waple, C. J., & Kawahara, T. (2009). Computer Assisted Language Learning system based on dynamic question generation and error prediction for automatic speech recognition. *Speech Communication*, 51(10), 995-1005. doi: <http://dx.doi.org/10.1016/j.specom.2009.03.006>
- Warschauer, M. (1996). Comparing face-to-face and electronic communication in the second language classroom. *CALICO Journal*, 13(2), 7-26.
- Watts, S., & Stenner, P. (2005a). Doing Q methodology: theory, method and interpretation. *Qualitative Research in Psychology*, 2, 67-91.
- Watts, S., & Stenner, P. (2005b, March). The subjective experience of partnership love: A Q Methodological study. *British Journal of Social Psychology*, 44(1), 85-107. doi: 10.1348/014466604X23473
- Watts, S., & Stenner, P. (2012). *Doing Q methodological research: Theory, method & interpretation*. Thousand Oaks, CA: Sage. doi: 10.1191/1478088705qp022oa
- Wehner, A. K., Gump, A. W., & Downey, S. (2011). The effects of Second Life on the motivation of undergraduate students learning a foreign language. *Computer Assisted Language Learning*, 24(3), 277-289.
- Williams, M., & Burden, R. L. (1997). *Psychology for Language Teachers*. Cambridge: Cambridge University Press.
- Winkel, S., & Petermann, F. (2007). Assessment of Achievement Motivation. In P. R. Zelick (Ed.), *Issues in the Psychology of Motivation* (p. 23-39). Nova.
- Yang, Y.-F. (2011). Engaging students in an online situated language learning environment. *Computer Assisted Language Learning*, 24(2), 181-198.
- Yashima, T. (2009). International posture and ideal L2 self in the Japanese EFL. In Z. D. . E. Ushioda (Ed.), *Motivation, language identity and the*

- L2 self* (p. 144-163). Multilingual Matters.
- Young, D. J. (1990). An Investigation of Students' Perspectives on Anxiety and Speaking. *Foreign Language Annals*, 23(6), 539-553.
- Zabala, A. (2014). qmethod: A Package to Explore Human Perspectives Using Q Methodology. *A peer-reviewed, open-access publication of the R Foundation for Statistical Computing*, 6(2), 163-173.
- Zabala, A., & Pascual, U. (2016). Bootstrapping Q Methodology to Improve the Understanding of Human Perspectives. *PLoS ONE*, 11(2). doi: 10.1371/journal.pone.0148087

Contributions

Chapters 2, 3, 4 and 5 are based on scientific manuscripts featuring the PhD candidate as first author. The manuscripts were co-authored with promotor Prof. dr. R. van Hout and co-promotors dr. C. Cucchiaroni and dr. H. Strik. For Chapters 3 and 4, authorship is also attributed to former PhD candidate and research team member dr. B. W. F. Penning de Vries. Dr. J. van Doremalen developed the ASR module. Additionally, significant work was completed by the PhD candidate developing the entire software platform and its components, including the logging capabilities. To allow for a full assessment of the PhD candidate's work, the contributions of the candidate and other members of the research team are described below.

Chapter 2

PhD candidate

The candidate surveyed the literature, analysed selected articles, and wrote the paper.

Co-authors

The co-promotors and promotor helped to define the scope of the paper, provided frequent feedback, and contributed suggestions and editing help during the writing process.

Chapter 3

PhD candidate

With the exception of the Automatic Speech Recognition (ASR) module,

the candidate implemented the entire software platform and its components, including the oral L2 grammar practice, in-practice questionnaire, logging system, proficiency tests, and background and post-practice questionnaires. He developed the approach for studying learners' affective experiences during practice, extracted data from the platform, analysed the data with guidance and assistance from the promotor, and wrote the paper.

Co-authors

The supervision team gave extensive feedback on the design of the oral L2 grammar practice and other components, and advised during the planning of the experiment. The promotor supervised the data analysis. During the writing process, the promotor and co-promotors pointed the candidate to relevant literature, and provided frequent comments, suggestions, and editing help. Dr. B. W. F. Penning de Vries provided extensive feedback on the design of the software components and frequent help testing the software. He and the PhD candidate contributed equally to organising and running the data collection.

Chapter 4

PhD candidate

As described above, the candidate implemented the entire software platform and its components, excluding the ASR module. He developed the approach for studying learners' affective experiences during practice, extracted data from the platform, analysed the data with guidance and assistance from the supervision team, and wrote the paper.

Co-authors

The promotor and co-promotors gave extensive feedback on the design of the experiments, and advised during planning phases. The promotor supervised the data analysis. During the writing process, the promotor and co-promotors pointed the candidate to relevant literature, and provided frequent comments and editing help. Dr. B. W. F. Penning de Vries provided extensive feedback on the design of the software components and frequent help testing the software. He and the PhD candidate contributed equally to organising and running the data collection.

Chapter 5

PhD candidate

The candidate designed and developed the computer application for eliciting ideal L2 selves, extracted the data from the system for analysis, performed the statistical analysis, and wrote the chapter.

Co-authors

The supervision team helped to plan the paper, reviewed the statistical analysis, and provided feedback and editing help during the writing process.

Samenvatting in het Nederlands

Het leren van een vreemde of tweede taal (T2) is een lange leerweg die ingegeven kan zijn vanuit uiteenlopende motieven. Taalleerders kunnen het intrinsiek leuk vinden, ze kunnen het doen om hun carrièreperspectieven te verbeteren, of ze hebben een rolmodel dat juist die T2 spreekt (Thompson & Vásquez, 2015). Verder dienen immigranten en vluchtelingen de taal van hun nieuwe land te leren als verplicht deel van de inburgering. Ongeacht waarom mensen een tweede taal leren en welke weg ze daarin bewandelen, alle leerders krijgen te maken met moeilijkheden en uitdagingen. In tegenstelling tot wat op internet met enige regelmaat gesuggereerd wordt, zullen de meeste leerders intensief en lang bezig zijn met het leren van de tweede taal, met perioden van vooruitgang, maar ook van stagnatie, en ze krijgen te maken met verschillende leersituaties. Onderzoek naar individuele verschillen in tweede-taalverwerving heeft zich gericht op de vraag waarom sommige leerders succesvoller zijn dan anderen (Ellis, 1994) en een belangrijk onderdeel daarvan is het onderzoek naar de rol van de persoonlijkheidskenmerken, motivatie, attitude, opvattingen, waarden, doelen, verwachtingen, percepties en emoties van leerders, gezamenlijk ook wel aangeduid als de affectieve kenmerken. De laatste twee decennia heeft dit soort onderzoek geleid tot een aantal belangwekkende theoretische en methodologische ontwikkelingen, inclusief, maar niet beperkt tot, een verhoogde aandacht voor het zelfbeeld van taalleerders, voor hoe affectieve kenmerken veranderen over de tijd en voor de rol van de context waarin een persoon leert en waarmee deze interacteert.

De afgelopen jaren zijn de computertechnologie en de taal- en spraaktechnologie een vanzelfsprekend deel geworden van taalleeromgevingen. Het leren van taal in computersettings staat bekend als Computer-Assisted Language Learning (CALL). CALL is een breed werkveld met een scala aan mogelijkheden (online leren, leren op afstand, mobiele toepassingen, applicaties voor zelfstudie op computers, interactieve systemen). CALL is ook een werkveld waarin diverse partijen betrokken zijn, zoals ontwikkelaars, leerders, docenten, onderzoekers en commerciële partners, elk vanuit hun eigen perspectief. De belangstelling voor CALL lijkt zich met name te richten op de praktische

inbedding van het taalleren, zoals het inbouwen van didactische principes en het design en minder op theoretische aspecten, alhoewel het belang van het onderzoek daarnaar wel wordt erkend (e.g. Levy, Hubbard, Stockwell, & Colpaert, 2015). Onderzoek naar CALL bestrijkt een breed scala aan onderwerpen, maar in feite gaat het om twee bredere gebieden: technologische innovaties aan de ene kant en didactische effectiviteit en inzicht in het leerproces aan de andere kant.

Het onderzoek waarover gerapporteerd wordt in dit proefschrift bestaat uit een combinatie van twee aandachtsgebieden: affectieve kenmerken die het taalleren beïnvloeden en onderzoek naar didactische effectiviteit en leerprocessen, bij het leren van taal in computerondersteunde leeromgevingen. We onderzoeken in het bijzonder affectieve ervaringen van een groep T2-leerders van het Nederlands tijdens hun taaltraining, in combinatie met factoren waarvan wordt verondersteld dat die deze ervaring kunnen beïnvloeden. De context is een computerapplicatie in de vorm van een zelfstudie die gesproken grammaticale training aanbiedt voor de Nederlandse grammatica. De gesproken grammaticale oefeningen betroffen de positie van het finiete werkwoord in de hoofdzin (zie hoofdstuk 3 en hoofdstuk 4). Het Nederlands is een taal met V2, Dat houdt in dat het finiete werkwoord ('verb', V), een syntactisch kenmerk dat voor de meeste T2-leerders nogal problematisch is.

Onze aanpak van het onderzoek kent twee perspectieven:

1. Een situatiegebonden perspectief waarin de focus ligt op de vraag hoe de affectieve kenmerken van leerders over de tijd veranderen en de relatie tussen affectieve kenmerken, oefengedrag en leeruitkomsten.
2. Een globaal perspectief waarin de inhoud en de diversiteit van het T2-gerelateerde toekomstige zelfbeeld van de leerders wordt onderzocht door gebruik te maken van een innovatieve onderzoekstechniek die via visuele, interactieve taken aangeboden wordt via het web.

Hieronder volgen samenvattingen van de onderzoekshoofdstukken in dit proefschrift, die afsluitend gevolgd worden door de belangrijkste conclusies.

Hoofdstuk 2: De evaluatie van de motivationele impact van CALL-systemen: huidig gebruik en toekomstige mogelijkheden

Hoofdstuk 2 geeft een overzicht van onderzoek naar affectieve dimensies in CALL. We hebben de huidige werkwijzen in de CALL-literatuur vergeleken met recente ontwikkelingen in de T2-motivatietheorie. Doel was om vast te stellen in hoeverre er overeenstemming of afstemming is en om suggesties te doen voor verder onderzoek. Hoofdstuk 2 omvat een survey van 22 studies waarin de volgende onderwerpen behandeld worden: het theoretische kader, de methode voor het verzamelen en analyseren van motivatiedata, de inbedding in het onderzoek van motivatie als een dynamische variabele, het gebruik van logboeken om het gedrag van leerders vast te leggen en de wijze waarop de motivatiekarakteristieken zijn onderzocht.

Onze voornaamste observaties met betrekking tot de huidige stand van zaken zijn de volgende:

1. ondanks het gegeven dat het merendeel van de studies situationele motivatie evalueert, zijn er weinig studies die een theoretisch kader gebruiken waarin motivatie wordt gezien vanuit een situationeel perspectief;
2. methoden om data te verzamelen balanceren tussen kwantitatieve metingen en kwalitatief exploratief onderzoek en een combinatie van deze methoden;
3. bij evaluaties van de situationele motivatie bestaat de neiging om slechts één sample te verzamelen na afronding van de training;
4. weinig studies maken gebruik van computerlogboeken om het leergedrag te bestuderen in relatie tot motivatie.

Op basis van deze resultaten doen we een aantal suggesties voor toekomstig CALL-motivatietoets. Onze eerste suggestie betreft de bevinding dat er erg weinig gespecialiseerde computerinstrumenten lijken te zijn om T2-motivatie te meten. CALL-onderzoekers lijken goed gepositioneerd om een bijdrage te leveren aan innovatieve computerinstrumenten die gericht zijn op het verzamelen van globale en situationele T2-motivatiedata. Een tweede suggestie is om meer gebruik te maken van logboekmodules, die vaak al aanwezig zijn in CALL-omgevingen. Logboeken kunnen in het bijzonder ingezet worden om te controleren of hetgeen leerders rapporteren over hun motivatie overeenkomt met wat ze daadwerkelijk doen en om oefeningen en patronen te identificeren die gerelateerd zijn aan motivatie. Onze derde suggestie is om een brug te slaan tussen globale motivationele kenmerken en theoretische kaders met betrekking tot tweede-taalverwerving. Als vierde suggereren we om te onderzoeken hoe globale en situationele motivationele data als bron gebruikt kunnen worden om instructies aan individuele leerders aan te passen aan de hand van CALL-systemen waarin leerdersmodellen worden gebruikt. Tot slot bevelen we met klem aan om kwalitatieve methoden te blijven gebruiken omdat deze belangrijke inzichten verschaffen in de globale en situationele motivatie van de leerder.

Hoofdstuk 3: Emotie-gebonden reacties van leerders in gecomputeriseerde T2-trainingen met correctieve feedback op gesproken grammaticale oefeningen

Leerders maken onvermijdelijk fouten tijdens het leren van een tweede of vreemde taal. Een belangrijke vraag is hoe nu met die fouten moet worden omgegaan. Correctieve feedback (CF), het corrigeren van fouten in de vorm van expliciete instructie, wordt als nuttig ervaren. Recente literatuur geeft aan dat CF een algemeen faciliterend effect heeft op de ontwikkeling van T2-vaardigheden, maar tegelijkertijd blijven er zorgen bestaan omdat

CF mogelijk een negatieve impact heeft op de emotie en motivatie van de leerder, met name in de context van het spreken in de T2. In hoofdstuk 3 is de impact van CF op emoties van de leerder onderzocht in een nieuwe context, namelijk het CALL (zelfstudie) systeem dat gesproken grammaticale oefeningen (over V2) aanbiedt. Hierbij wordt Automatische SpraakHerkenning (ASH) technologie ingezet voor automatische CF wanneer er fouten in de woordvolgorde optreden. In een experiment hebben we twee versies van het CALL-systeem getest: één CF versie maakte gebruik van ASH voor onmiddellijke feedback. In een tweede versie zonder feedback (NOCF), kregen de leerders slechts te horen dat de opnames waren opgeslagen en dat ze het opnieuw konden proberen of door konden gaan naar de volgende vraag. Deelnemers oefenden op twee verschillende dagen met het systeem. Tijdens het oefenen volgden we de emotionele ervaringen van de leerders door gebruik te maken van een module waarin leerders periodiek werd gevraagd om hun gevoelens te rapporteren (plezier, bereidheid om te oefenen en zelfeffectiviteit) middels een korte vragenlijst. Bovendien werd de inspanning die het de leerder kostte en de werkwijze bijgehouden in logbestanden van de opnames.

De analyse van de gerapporteerde gevoelens leverde geen significante verschillen op tussen de NOCF- en CF-conditie, hetgeen betekent dat correctieve feedback geen negatief effect had op motivatie en de andere affectieve kenmerken. De resultaten laten bovendien zien dat emotie kan veranderen tijdens de training van de T2. Aan het einde van de training leken leerders bijvoorbeeld significant minder bereid om te oefenen en rapporteerden significant meer zelfeffectiviteit in het beantwoorden van de vragen. Een vergelijking tussen de gerapporteerde gevoelens tijdens de eerste en de laatste oefensessie laat zien dat het oefenen als plezierig werd ervaren en dat de leerders meer bereid waren om te oefenen en meer zelfeffectiviteit rapporteerden. Analyses van de gedragsscores lieten een significant verschil zien tussen de NOCF- en de CF-conditie. Leerders in de NOCF-conditie waren geneigd om zich tot één poging te beperken om een vraag te beantwoorden. In de CF-conditie, waarin de leerders werd verzocht om hun fouten te corrigeren voordat ze verder konden gaan, deden zij meerdere pogingen per vraag (ze deden dus meer moeite). Een analyse van de correlaties tussen gerapporteerde gevoelens en gedragsscores liet een significant positieve relatie zien tussen gedragsscores en zelfeffectiviteit, maar dit was alleen het geval voor de CF-conditie.

De resultaten van hoofdstuk 3 laten zien dat CF de leerders stimuleert om meer gesproken uitingen te produceren zonder dat dit de affectieve dimensies van het leerproces negatief beïnvloedt, dat gesproken grammaticale oefeningen de leerders een positieve leerervaring geven die aanhoudt na herhaaldelijk oefenen, en dat de leerders CF zouden kunnen gebruiken om hun eigen functioneren te monitoren en om de inschatting van hun zelfeffectiviteit bij te stellen.

Hoofdstuk 4: Zelfstudie op basis van computergebruik voor het leren van gesproken grammaticale oefeningen: emotionele ervaringen, correctieve feedback, vaardigheid en leerwinst

Hoofdstuk 4 presenteert een onderzoek naar de evaluatie van mondelinge grammaticale T2-oefeningen die we hebben ontwikkeld voor computergebruik. De oefeningen betroffen het verschijnsel van V2 in het Nederlands. Naast de impact van CF en de gerapporteerde gevoelens tijdens de trainingssessies hebben we leeruitkomsten en individuele leerdersverschillen in de analyses betrokken.

Aan de hand van taalvaardigheidstesten zijn de correctheidsscores voor de plaatsing van het finiete werkwoord voor en na de trainingssessies vastgesteld. Op basis van deze correctheidsscores hebben we de leerders vervolgens een label toegekend waarmee we hun aanvankelijke V2-vaardigheid (laag of gemiddeld) en leeruitkomsten (vooruitgang of geen vooruitgang) hebben gekarakteriseerd. Deze labels zijn daarna als onafhankelijke variabelen gebruikt om te onderzoeken wat het effect is van het niveau van vaardigheid en leeruitkomsten op de mate van plezier bij het oefenen, de bereidheid tot oefenen en de zelfeffectiviteit die door leerders is gerapporteerd tijdens de eerste en de laatste trainingssessies. Aan de hand van een vragenlijst die na de oefensessies werd ingevuld, is een analyse gedaan van de meer globale houding van de leerders ten opzichte van het systeem.

Het is opmerkelijk dat CF op zich weinig impact had. Er werden geen significantie verschillen gevonden tussen de NOCF- en de CF-conditie. De gerapporteerde gevoelens van de leerders waren hetzelfde tijdens de eerste en de tweede sessie; de resultaten lieten zien dat het oefenen als plezierig werd ervaren door de leerders, dat ze bereid waren om te oefenen en dat de moeilijkheidsgraad adequaat was. De analyse waarin initiële vaardigheid en leerwinst waren opgenomen liet zien dat V2-leerders die initieel gemiddeld vaardig waren veranderingen in gerapporteerde gevoelens hadden die parallel liepen met de veranderingen in hun V2-vaardigheid. Gemiddeld vaardige leerders die vooruitgang boekten, rapporteerden positieve ervaringen tijdens de eerste en de tweede sessies, en gemiddeld vaardige leerders die geen vooruitgang boekten, waren minder positief tijdens de laatste sessie. Minder vaardige leerders rapporteerden daarentegen positieve ervaringen in de eerste en tweede sessies, ongeacht of er vooruitgang was of niet. Uit de resultaten van de vragenlijst die na de laatste oefensessie werd afgenomen, bleek dat minder vaardige leerders die feedback kregen tijdens het oefenen, de oefeningen aantrekkelijker en gemakkelijker vonden.

Samenvattend bleek dat CF geen significante, (negatieve) invloed heeft op de affectieve kenmerken van leerders, maar dat individuele verschillen zoals initiële vaardigheid en leeruitkomsten wel een rol spelen. De feedback op woordvolgorde (V2) wordt met name door minder vaardige leerders op prijs gesteld.

Hoofdstuk 5: Onderzoek naar motivationele zelfbeelden met behulp van ‘possible selves’: een computerapplicatie voor volwassen T2-leerders van het Nederlands

Hoofdstuk 5 begint met de aanname dat leerders kunnen verschillen in motivatie op de lange termijn (Csizér & Dörnyei, 2005a; Papi & Teimouri, 2014). Deze verschillen kunnen een rol spelen bij emotionele ervaringen of gedrag tijdens T2-training (Dörnyei, 2002; Julkunen, 2001). We presenteren een nieuwe computerapplicatie die is ontwikkeld om te kunnen onderzoeken wat de ideale T2-selves zijn van een groep volwassen Nederlandse T2-leerders. Deze groep heeft meegedaan aan de experimenten die zijn beschreven in hoofdstuk 3 en 4. Aan de deelnemers werd gevraagd om verschillende toekomstbeelden van zichzelf (‘futures’) te evalueren. Aan de hand van de computerapplicatie bladerden, selecteerden en sorteerden zij kaarten waarop verschillende mogelijke futures waren beschreven (de ‘possible selves’). De beoordelingen van de wenselijkheid van verschillende futures werden geanalyseerd met een Q-analyse (een geïnverteerde factoranalyse) om de aard en diversiteit van ideale T2-zelfbeelden te onderzoeken. Uit de analyse van de gesorteerde data kwamen 14 ideale T2-zelfbeelden naar voren, waarvan de drie kernprofielen in detail zijn geanalyseerd, omdat deze veruit het meest voorkwamen. Ondanks dat de drie profielen verschilden in hun prioriteiten, bijvoorbeeld in het belang dat ze hechten aan mogelijke carrière of andere pragmatische voordelen van het leren van het Nederlands, bleek dat alle drie profielen een gemeenschappelijke noemer hadden, namelijk futures die gerelateerd zijn aan het spreken van T2. In combinatie met de bevindingen uit hoofdstuk 3 en 4 is het voorstelbaar dat de positieve reacties van de leerders op het systeem vooral te maken hebben met het feit dat het systeem het mogelijk maakt mondelinge taalvaardigheid te oefenen, hetgeen overeenkomt met hun wens om in de toekomst een T2 te kunnen spreken. De resultaten van hoofdstuk 5 laten ook zien dat de deelnemers in ons onderzoek verschillende ideale T2-zelfbeelden hadden, wat suggereert dat leerders zouden kunnen profiteren van geïndividualiseerde leersystemen die uitdrukkelijk rekening houden met motivatiekenmerken. De methode die in dit hoofdstuk werd voorgesteld biedt een concreet handvat om die motivatiekenmerken te achterhalen.

Conclusies

Dit proefschrift verbindt nieuwe mogelijkheden van taal- en spraaktechnologie met ontwikkelingen in taalleer- en motivatietheorieën om zodoende een meer omvattend taalleerplatform te ontwikkelen. Doel is uiteraard een beter leerplatform, maar tegelijkertijd biedt de ontwikkeling van een dergelijk platform mogelijkheden om belangrijke theoretische kwesties rondom aard en functie van gesproken grammaticale oefeningen in de tweede taal te bestuderen en te beslechten. De kennis en de informatie die voortkomen uit ons onderzoek laten zien dat leeromgevingen waarin gebruik wordt gemaakt van computers

waardevolle handvatten bieden voor verder onderzoek. Deze leeromgevingen kunnen in potentie aanzienlijk bijdragen in het doorgronden van de rol en functie van affectieve factoren en in een verdere verfijning en individuele afstemming van het didactisch design voor ten behoeve van het leren van een tweede taal.

Curriculum Vitae

Stephen Bodnar was born in 1978, in Richmond Hill, Canada. He obtained a Bachelor's degree from the University of Alberta in 2002, specialising in computing science. Following graduation, he was employed in the software industry and as an English teacher in South Korea. After obtaining his Master's degree from Macquarie University in speech and language processing in 2009, he moved to Los Angeles to work as a scientific programmer with Alelo Ltd. In 2010, he relocated to Nijmegen to start his PhD as part of the research project Feedback and the Acquisition of Syntax in Oral Proficiency.

At present, he is employed as a research assistant by the ATLAS Lab at McGill University, where he will begin a postdoctoral fellowship in the summer of 2016. His current work involves collaborating with a multi-disciplinary team of researchers to develop and evaluate technology-rich learning environments